UNIVERSITY COLLEGE LONDON
DEPARTMENT OF PHYSICS AND ASTRONOMY

Computational studies of electron transfer in the bacterial deca-heme cytochrome MtrF

Marian Breuer

Supervisor: Dr. Jochen Blumberger

Sponsors:
University College London
Pacific Northwest National Laboratory

A thesis submitted for the degree of
Doctor of Philosophy from University College London.

July 23, 2015

To whomever is reading this thesis

Declaration

I, Marian Breuer, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Certain bacteria possess the remarkable ability to respire on extracellular solid metal oxides, which requires a unique biomolecular machinery optimized for long-range electron transport. To perform this function efficiently, microorganisms have adapted multi-heme c-type cytochromes to arrange heme cofactors into wires that cooperatively span the cellular envelope, transmitting electrons along distances greater than 100 Å before they are passed on to extracellular terminal acceptors via direct contact or secreted soluble electron shuttles (like flavins). While the interactions and "whole-protein" properties of these special multi-heme proteins have been the subject of intensive research efforts, molecular-level insight has long been elusive.

Here, a wide range of computational methods is deployed to study the electron transfer (ET) properties of outer membrane (OM)-associated deca-heme cytochromes in the bacterium *S. oneidensis*, from explicit electronic structure calculations to empirical ligand docking. In particular, thermodynamic and kinetic parameters for heme-to-heme ET in the OM deca-heme cytochrome MtrF are calculated which enables us to model through-protein steady-state electron transport. Furthermore, the interaction of its homologue MtrC with the soluble shuttle flavin mononucleotide is studied via docking simulations.

Our calculations of heme redox potentials in MtrF yielded a free energy surface with two "up-hill" steps of 0.2 eV for through-protein electron transport. These potential kinetic obstacles were found to be counteracted by stronger electronic interactions precisely between those cofactors that form the free energy hills. This correlation of thermodynamic properties and electronic interactions was found to be essential for the cytochrome to yield through-protein electron transport rates of 10^4-$10^5\,s^{-1}$ (as determined from our steady-state electron transport calculations), consistent with an experimentally established lower limit.

Our docking studies reproduce an experimentally determined affinity quite well and are consistent with the existing hypothesis regarding the role of flavin molecules as electron shuttles; they furthermore suggest concrete binding sites on the deca-heme cytochrome MtrC.

Overall, by deploying and combining a range of computational approaches, this work has yielded molecular-level insights into OM deca-heme cytochromes that are difficult to obtain from experiment.

Acknowledgments

First, I would like to thank my supervisor Dr. Jochen Blumberger for all his help and support. He was always there to help me with my project no matter how long a particular topic discussion would take and always radiated a good working atmosphere. I am also grateful to Dr. Kevin Rosso, a collaborator at Pacific Northwest National Laboratory, for his support. I also would like to thank my family back home for their support in many ways. I am also grateful to my past and present colleagues: Dr. Po-hung Wang, Dr. Varomyalin Tipmanee, Dr. Harald Oberhofer, Dr. Martina Mangold, Dr. Adam Kubas, Fruzsina Gajdos, Jacob Spencer, Guido Falk von Rudorff, Karina Chan, Hui Yang, Dr. Ehesan Ali, Bastian Burger, Benjamin Rosseau, Siim Valner, Felix Hoffmann, Felix Hummel, Jamie Sage, Marco Montefiori and William Ashworth.

This work was made possible by an IMPACT studentship jointly sponsored by University College London and Pacific Northwest National Laboratory.

This work made use of the facilities of HECToR and ARCHER, the UK's national high-performance computing services, which were/are provided by UoE HPCx Ltd at the University of Edinburgh, Cray Inc and NAG Ltd, and funded by the Office of Science and Technology through EPSRC's High End Computing Programme. A portion of the research was performed using EMSL, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory. The use of the UCL Legion High Performance Computing Facility, and associated support services, in the completion of this work is also acknowledged.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Outer-membrane cytochromes in the respiration of extracellular substrates

Energy transduction in biological systems very often depends on rapid and reliable transfer of electrons over distances on the nanometer scale [14]. While electron exchange between Fe^{2+} and Fe^{3+} ions in solution occurs on a time scale of seconds [15], the time scale is reduced to microseconds (or less) when the ions are embedded in cofactors in a protein [16, 17]. A particular electron transport phenomenon of intense current research interest is the extraordinary ability of certain Gram-negative bacteria to utilize extracellular insoluble transition metal oxides as terminal electron acceptors in their respiratory chain [18], reducing minerals like ferrihydrite or manganese oxide [19]. This is quite unusual as most other terminal electron acceptors used by life can be imported into the cell to be reduced (e. g. O_2). Respiration of extracellular substrates however requires an efficient system to transport electrons from the cytoplasmic (inner) membrane (IM), where they usually accumulate from the oxidation of organic substrates, across the bacterium's periplasm (inter-membrane space) and outer membrane (OM) to the substrate [20]. The reduction of insoluble metal oxides through microorganisms is an important biogeochemical process that is far from being understood so far but plays an important role in the geochemical cycling of different metals [18]. In addition to shedding light on these important environmental processes, the study of these electron transfer (ET) systems also holds out the prospect of promising technological applications like biofuel cells, which could couple the breakdown of biological waste to the generation of electricity [21], and modification of other microorganisms to combine electron transfer abilities with other useful properties like photocatalytic activity [22].

A representative of these dissimilatory metal reducing bacteria (DMRB)

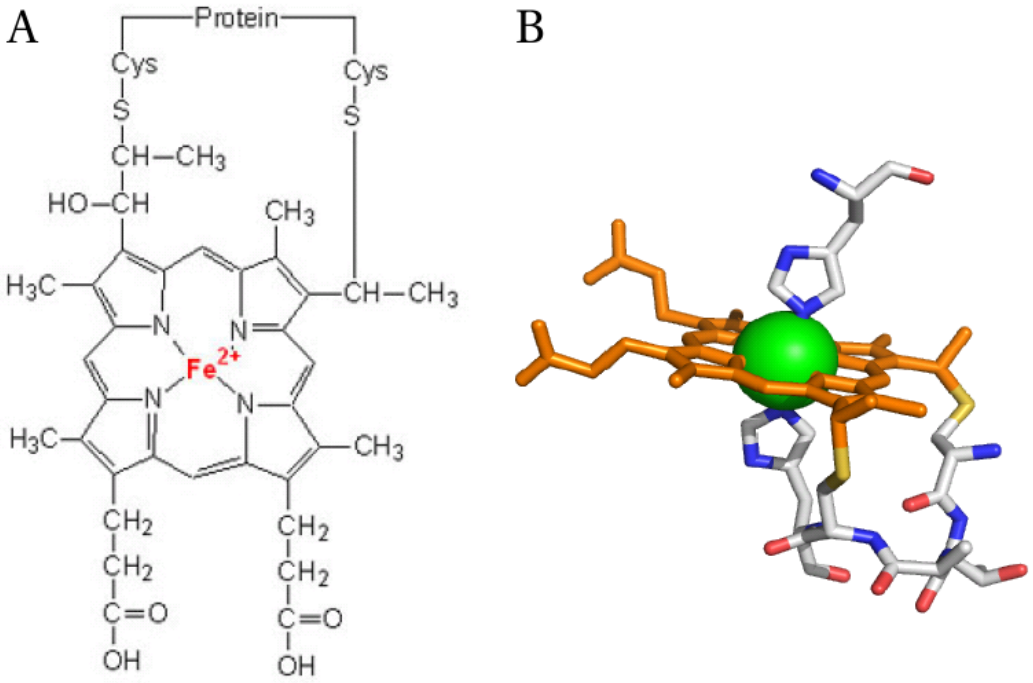

Figure 1.1: Molecular structure of heme c. (A) Lewis structure showing the central iron ion and the aromatic porphyrin ring surrounding it (coordinating to the iron via four nitrogens). The cysteines linking the porphyrin ring to the protein are abbreviated as "Cys". The axial ligands are omitted for clarity. (B) Three-dimensional structure showing the heme in orange (iron as a green sphere) and the protein residues coloured by atom (carbon: gray, nitrogen: blue, oxygen: red, sulphur: yellow.) The iron is axially coordinated via two histidine side chains, the lower of which is part of a CXXCH binding motif (cysteine-two arbitrary residues-cysteine-histidine; see text). Adapted from Ref. [1] with permission from the Royal Society of Chemistry. The original figure had been prepared by Dr. Julea Butt.

that has been studied very extensively so far is *Shewanella oneidensis* strain MR-1, a facultatively anaerobic bacterium discovered first in Lake Oneida, NY, USA in 1988 [19, 23]. The interaction of *S. oneidensis* with extracellular substrates has been studied in great detail and has been found to depend on a system of cytochromes, i. e. proteins containing heme cofactors. A heme consists of an iron ion (most often in formal oxidation state II or III) ligated by a porphyrin, a planar organic moiety with four nitrogens surrounding and coordinating the central iron, see Fig. 1.1. This equatorial coordination then allows for up to two additional axial ligands which in the cytochromes in *S. oneidensis* are usually two histidine side chains from the protein. Furthermore, hemes in cytochromes usually also bear two covalent links to the protein via its side chains forming thioether bonds with cysteine residues (this type of heme is called heme c); these cysteines, together with one of the histidines, form a specific binding motif ('CXXCH', where C stands for cysteine, H for histidine and X for an arbitrary amino acid). The other func-

tionality of interest are the two acidic propionate side chains of the heme, giving it a negative total charge if both are deprotonated. Hemes can change easily between an oxidized state (Fe(III)) and a reduced state (Fe(II)), rendering them versatile cofactors for transport of electrons within a cellular environment. (Although the entire heme loses/gains electron density during oxidation/reduction, the iron's formal oxidation state is used here to denote the redox state of the cofactor.) *S. oneidensis* employs a special group of cytochromes: Multi-heme cytochromes, defined as cytochromes with hemes brought together in close distance (Fe-Fe distance 15.5 Å or smaller) [24, 25]. This close packing of hemes (sometimes ten or more in one protein) allows rapid transport of electrons along the cofactor chain.

Multi-heme cytochromes are at the core of the respiratory capabilities of *S. oneidensis*. Fig. 1.2 depicts the extracellular respiration of *S. oneidensis* schematically: organic substrates are broken down (oxidized) in the cytoplasm and their electrons transported from the inner membrane (IM) through the periplasm and outer membrane (OM) to multi-heme cytochromes located on the cell exterior. The final ET from these OM cytochromes to an extracellular insoluble substrate has been proposed to take place both via direct contact (b in Fig. 1.2) [21, 26, 10] and indirectly via soluble redox shuttles, in particular flavins (a in Fig. 1.2; 'F' for 'flavin'). [27, 28] (Flavins seem also capable of chelating Fe(III) ions in solution, thus enhancing their availability, and coating mineral oxide surfaces, possibly altering their reactivity [29].) However, the relative physiological importance of these mechanisms is still unclear; both direct and indirect electron transfer might even take place under the same conditions [30]. In addition, the extracellular ET (EET) can take place directly at the cell surface or via μm-long conductive appendages [31, 32], often termed bacterial nano-wires. (Indicated in Fig. 1.2 as the bud in c.) These nano-wires were shown to exhibit p-type semiconductor behaviour [33] and were recently found to be extensions of the outer membrane containing OM cytochromes [34], which might mean that they are able to reduce insoluble substrates along their entire length (directly and/or indirectly).

The nature of the multi-heme cytochromes on the cell surface, and the ones that transport electrons to them from the IM, has been extensively studied and considerable insight into the involved proteins, their properties and interactions has already been obtained. The currently known multi-heme cytochromes of *S. oneidensis* (structures, locations and substrates) are shown in Fig. 1.3. The electron transport network begins at the IM where substrate oxidation yields quinols (QH_2) which is oxidized by the membrane-bound tetraheme cytochrome CymA. CymA is able to pass its electron on to, amongst others, the deca-heme cytochrome MtrA [28]. MtrA can form a complex

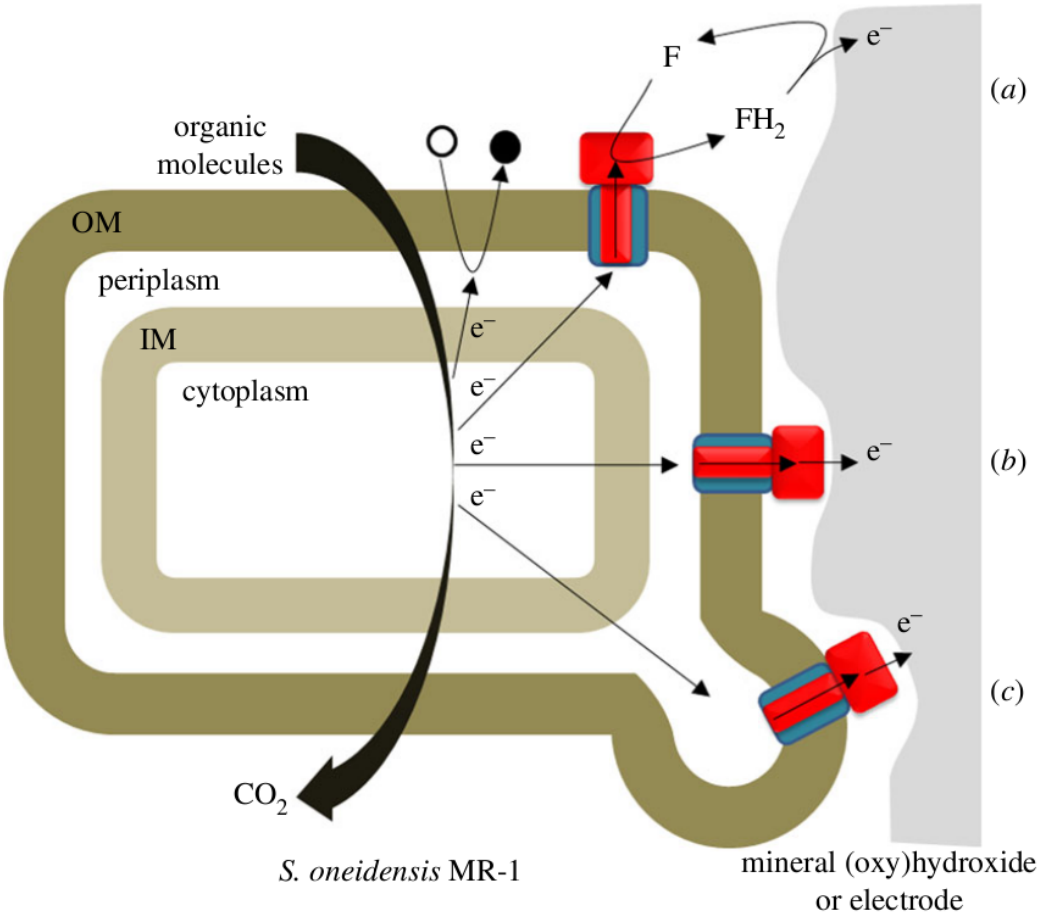

Figure 1.2: Different mechanisms for extracellular ET in *S. oneidensis*. The oxidation of organic molecules releases electrons to the inner membrane (IM). From the IM electrons can be passed on to soluble terminal electron acceptors (black circle) that enter the periplasm; alternatively, they can be transported across the periplasm and outer membrane (OM) to extracellular terminal electron acceptors. ET from the cell surface may be mediated by redox shuttles, particularly flavins (F) (a), occur via direct contact with an OM cytochrome (b), or involve cellular appendages called nano-wires (c). The latter are OM extensions (represented by the bud in (c)) that contain OM cytochromes and can reach lengths on the μm-scale. Transmembrane cytochrome complexes are represented by a blue rectangle with a red stripe and red rectangle (the latter representing the OM cytochrome). Reproduced from Ref. [1] with permission from the Royal Society of Chemistry (figure prepared by Dr. Julea Butt).

with the OM β-barrel protein MtrB which on its exterior side binds the deca-heme cytochrome MtrC; MtrB is in this context believed to act as a pore in which MtrA and MtrC can come sufficiently close to exchange electrons [20]. MtrC is one of the OM cytochromes central to the reduction of extracellular solid substrates: It has been shown to reduce solid substrates directly [20, 10] and has been proposed as the major reduction site for flavin mononucleotide (FMN) [35], a redox shuttle secreted by *S. oneidensis* [29]. It is this di-cytochrome-transmembrane complex that is at the core of *S. oneidensis*'

capability to respire on extracellular substrates: In fact, it has been shown that the MtrA-MtrB-MtrC system (or short, MtrCAB) can be transferred to other microorganisms as well and is sufficient to significantly enhance their ability to reduce insoluble haematite [22].

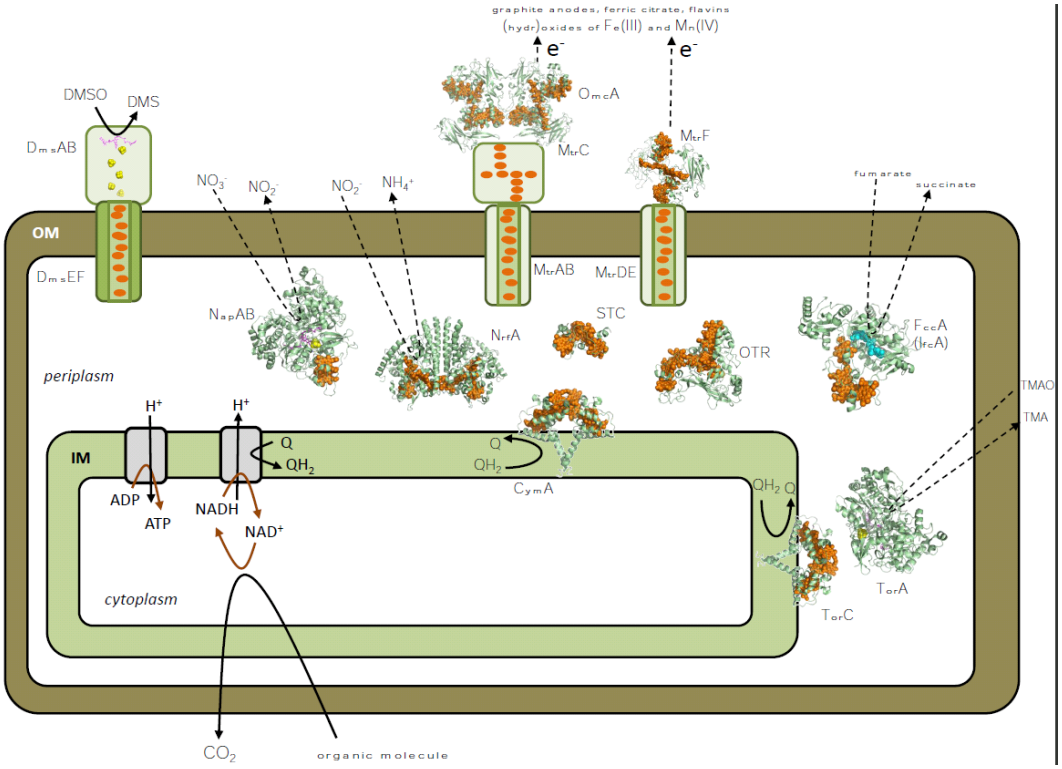

Figure 1.3: Multi-heme cytochromes from *S. oneidensis* illustrated schematically to indicate their cellular location and roles. Structures are shown for proteins from *S. oneidensis* STC (pdb code 2K3V), NrfA (3UBR), OTR (1SP3), FccA (1QJD), MtrF (3PMQ) and OmcA (4LMH); and their homologues NapAB (1OGY), CymA (2J7A), TorC (2J7A) and TorA (1TMO). The arrangement of cofactors in DmsAB is based on that in the homologue NarGH (1R27). At the inner membrane (IM), the respiratory processes are depicted that generate adenosine triphospate (ATP) and pass electrons to quinones (Q) in the IM. Hemes are shown in orange, FeS clusters in yellow/green, flavin adenine dinucleotide (FAD) in cyan and molybdopterin in purple. Reproduced from Ref. [1] with permission from the Royal Society of Chemistry (figure prepared by Dr. Julea Butt).

MtrC has two homologues in *S. oneidensis*: OmcA, which associates with MtrC in a 2:1 stochiometry [36, 37, 38] (see complex model in Fig. 1.3); and MtrF, whose gene is part of an operon containing genes for MtrD (a homologue of MtrA) and MtrE (a homologue of MtrB). It is thus hypothesized that MtrD, MtrE and MtrF form a transmembrane ET complex MtrDEF analogous to MtrCAB. [39] The precise scopes of the functions of MtrC and its two homologues OmcA and MtrF are still not entirely clear. On the one hand, some specificities have been observed: While MtrC has been proposed to play the major role for reduction of FMN, OmcA is important in the reduction of the

related riboflavin (RF) [35]. MtrF (together with MtrD and MtrE) is most highly expressed during biofilm growth [40]. On the other hand, some modularity could be established as well: Hybrid complexes can form from MtrCAB and MtrFDE components [39] and it has also been shown that while MtrF does not contribute to iron oxide reduction in wild type cells, it is apparently able to assume the role of MtrC in iron oxide reduction if the gene *mtrC* is deleted; however, this activity still depends on MtrA and MtrB, and while some natural substrates for MtrC could already be established (see above), the actual substrate for the assumed MtrD-MtrE-MtrF pathway remains unknown [39].

Both the properties of individual multi-heme cytochromes as well as their interactions amongst each other have been studied quite extensively so far. Electrochemical properties of individual cytochromes are routinely studied via cyclic voltammetry [20, 2, 41] which yields the overall response of a cytochrome to an applied bias potential; sometimes, this can be used to obtain individual macroscopic redox potentials [2]. Interaction with substrates involves studies of substrate binding [21, 42, 13] and substrate reduction rates [2, 43]. *In vitro* and *in vivo* studies have also established interactions between specific multi-heme cytochromes [36, 37, 38, 44]. However, insight into the large OM cytochromes (which are of particular interest as the interface between microbe and mineral) has been hampered in the past by the lack of structural information, with the studies mentioned above only yielding "whole-protein" level information for these cytochromes. E. g., while the redox potential range of operation for MtrC could be detected from voltammetry [45], only speculations could be made in regard to the protein's need for ten cofactors. This constitutes a significant obstacle in elucidating the exact functions of these cytochromes. Molecular-level insight is also necessary to understand and extract design principles for these complex biomolecules which would not only promote understanding of the natural system but might even provide concepts useful in the design of artificial proteins.

This situation changed a few years ago with a seminal paper presenting the first crystal structure of an OM deca-heme cytochrome from *S. oneidensis*: MtrF, the homologue of MtrC [2]. The most remarkable feature of this structure (which is discussed in detail in Section 1.2) is its nonlinear arrangement of hemes, with eight hemes forming a central heme chain traversing the protein and the remaining two hemes branching off in the middle, thus yielding a two-dimensional topology of cofactors. Recently, the structures of OmcA and MtrC could be resolved as well [46, 3] and revealed the very same heme arrangement. (Another homologue, the undeca-heme cytochrome UndA from *Shewanella sp*. Strain HRCR-6, was also found to feature the same heme ar-

rangement, plus an additional eleventh heme. [43]) An available structure not only allows one to form new hypotheses based on the structure alone (e. g. possible substrate binding regions [2]), but also enables one to study these proteins with computational methods - yielding molecular-level insight difficult to obtain in experiment. Given that these cytochromes serve to transport electrons, the intraprotein ET properties are of course of major interest; other properties that would be both interesting and possibly attainable would be interactions with substrates or other proteins.

This thesis focuses on the computational study of OM cytochromes in *S. oneidensis* (mainly MtrF). In the work presented in this thesis, thermodynamic and kinetic parameters of ET between heme cofactors in MtrF were calculated using a variety of computational methods; these parameters for heme-to-heme ET were then used to study electron flow through the entire protein. Finally the interaction with soluble substrates was investigated, specifically the binding of FMN to MtrC. While the calculations on thermodynamic properties of MtrF presented in this thesis [6] were the first computational study on MtrF to be published to the author's best knowledge, other authors have since then also deployed computational approaches to study different functional aspects of this deca-heme cytochrome. Smith and Rosso [47] investigated possible gating effects for electron conduction through MtrF; and Byun et al. [48] used the heme-to-heme ET rate constants from this work [4] to model electron flow through MtrF with a Kinetic Monte Carlo (KMC) algorithm. Nakano et al. [49] extended these KMC simulations to a modeled MtrF-OmcA complex (obtained from protein docking simulations).

As the structure of MtrF is the basis for all simulations presented in this work, it is discussed in detail in the next section - together with the structure of MtrC which is the focus of the substrate-binding studies in the last part of this work. Subsequently, a detailed overview of this thesis is given in Section 1.3.

1.2 The crystal structures of MtrF and MtrC

The crystal structure of MtrF has for the first time been described by Clarke et al. [2]. The protein consists of 593 amino acid residues arranged in four domains (see figure 1.4); two of these domains (referred to as domain I and domain III, based on the order of appearance in the amino acid sequence) contain an extended β-sheet each, while the other two domains (II and IV) contain several α-helices and covalently bind five heme groups per domain via CXXCH binding motifs (see discussion in previous section). The iron atom is ligated by two histidine residues in each cofactor, one of the his-

tidines belonging to the binding motif. The spatial arrangement of the hemes is similar between domains II and IV; both domains contain three stacked, roughly parallel cofactors (3,4,5 and 8,9,10, respectively) and two approximately coplanar heme groups (1,2 and 6,7, respectively) facing the triple-stack (with porphyrin planes orthogonal to to planes in the stack). In the overall protein structure the two domains are assembled such that the coplanar heme pairs form a line of four roughly coplanar cofactors through the central region of the protein, with the two outer hemes (2 and 7) facing domains I and III, respectively, which flank the two central domains II and IV. The triple-stacks branch off this central line in opposite directions, giving rise to a crossed arrangement of cofactors where the six stacked cofactors together with the two central coplanar ones form a continuous octa-heme chain, while the outer two of the four coplanar hemes form side branches of this major heme chain. This motif has been termed "staggered cross" by Clarke et al. Individual pairs of neighbouring cofactors can be grouped according to one of the three orientation motifs: The pairs 1-2, 1-6 and 6-7 show a (roughly) coplanar arrangement; 1-3 and 6-8 feature hemes with approximately orthogonal porphyrin planes and are thus referred to as "T-shaped" in the following; 3-4, 4-5, 8-9 and 9-10 finally consist of hemes with stacked porphyring rings. (To be clear, "stacked" denotes partial stacking of the porphyrin rings - the axial ligands would of course prevent the complete stacking of these hemes.)

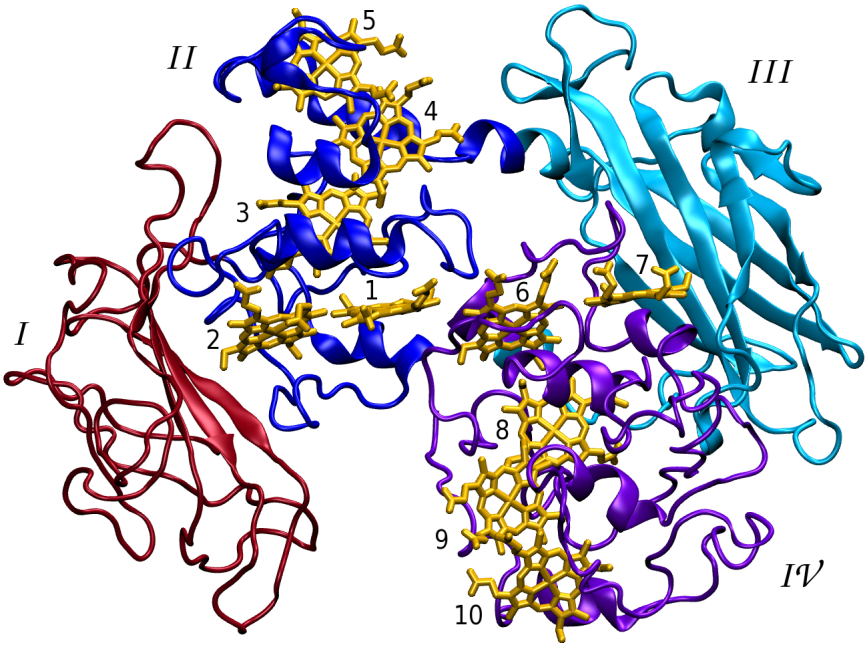

Figure 1.4: Crystal structure of MtrF (pdb code: 3PMQ) [2]. Roman numerals refer to different domains (shown with backbone only and secondary structure elements) while Arabic numerals denote the different heme cofactors (shown in yellow).

Clarke et al. hypothesize that domains I and IV serve to form contacts

with other proteins, and that the heme cofactor at the end of the octa-heme chain in domain IV (heme 10) could act as the electron entrance site in MtrF; the cofactor at the opposite end of the heme chain in domain II (i.e. heme 5) could then serve as an electron egress site towards mineral oxides as its configuration has previously been found to be optimal for this electron transfer in a related model compound [26]. It also might be used to transfer electrons to other OM cytochromes.

As mentioned in the previous section, the structures of homologues of MtrF could be resolved since this first structure was published [43, 46, 3]. The homologue structures revealed the same staggered cross of hemes. Fig. 1.5A shows the structure of MtrC, the cytochrome whose interaction with the soluble shuttle FMN is studied in Chapter 5. As can be seen, its overall features are very similar to MtrF. (Panel B depicts the molecular structure of FMN.)

1.3 Thesis statement

The aim of this work is to provide detailed insight into the properties of OM cytochromes of *S. oneidensis*, specifically MtrF and MtrC, on a molecular level using computational modelling methods. These allow us to address questions that are difficult to tackle by experiment. Specifically, thermodynamic and kinetic parameters for electron transfer between the ten cofactors in MtrF are calculated, and hence for transport through the entire protein, starting from the atomistic structure of MtrF. Using the framework of nonadiabatic Marcus theory allows us to understand electron transfer in terms of driving forces, free energies of (protein and solvent) reorganization and electronic couplings between heme cofactors. The results not only allow for predictions in regard to overall trans-protein transport rates and hypotheses regarding the function of different parts of the protein but also enable to draw correlations between protein structure and function and thus to deduce design principles. Fig. 1.6 illustrates the multi-scale nature of this topic: To describe electron transport through MtrF as a whole (A), ET rates are needed for pairs of adjacent hemes (B) which requires calculation of the electronic structure of individual hemes and averaging of heme properties over the thermally fluctuating protein conformations (C).

For MtrC we investigate the interaction with soluble flavins, in doing so also trying to resolve properties on a molecular level so far not attainable in experiment. Specifically, we use the atomistic structure to study where and how strongly FMN could bind to the surface of MtrC.

This thesis is organized as follows. Chapter 2 presents the theoretical foundations of the methods utilized in this work, covering: Semi-classical

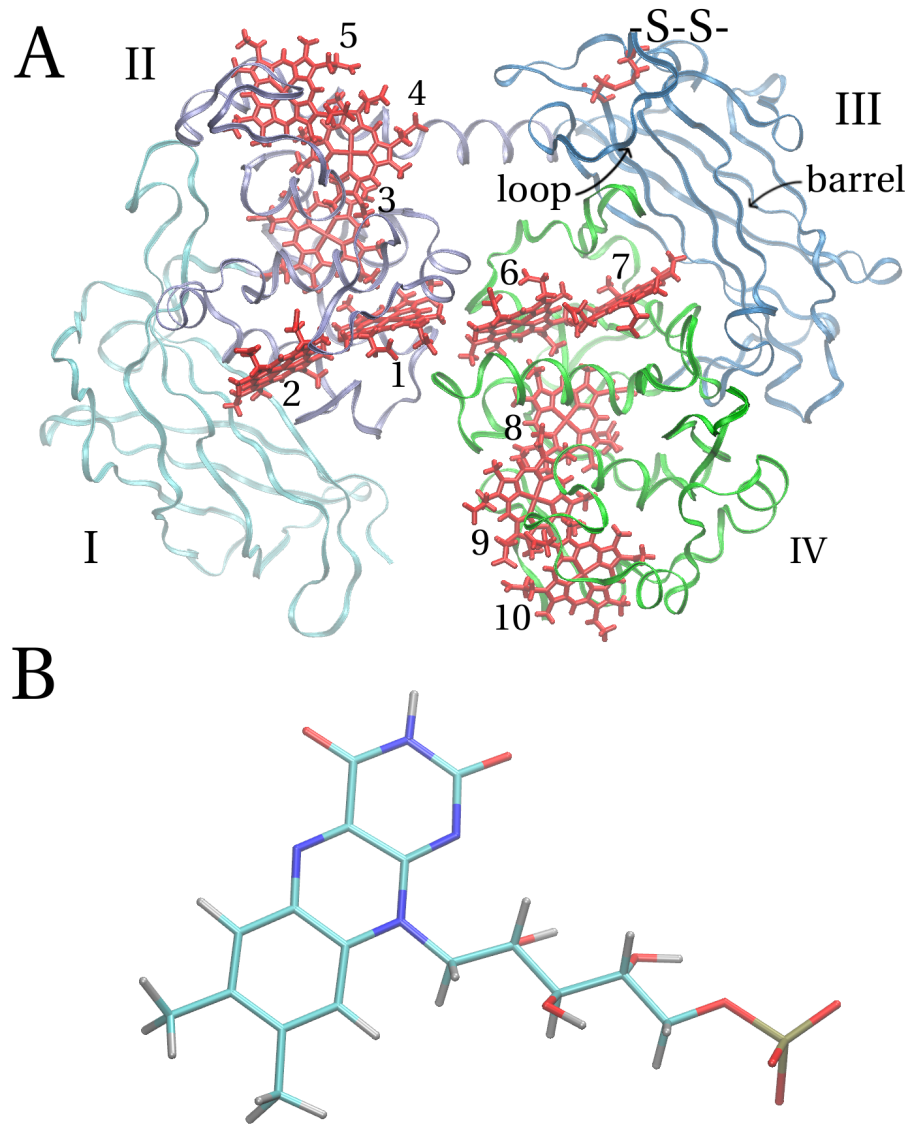

Figure 1.5: (A) Crystal structure of MtrC (pdb code 4LM8) [3]. Roman numerals refer to different domains (shown with backbone only) while Arabic numerals denote the different heme cofactors (shown in red). The barrel and a disulphide bond in Domain III are labelled and are further discussed in Chapter 5. (B) The molecular structure of flavin mononucleotide (FMN), whose binding to MtrC is studied in Chapter 5. Carbon is shown in cyan, nitrogen in blue, oxygen in red and hydrogen in silver. The tricyclic head group is the redox-active moiety.

nonadiabatic Marcus theory which is used to describe heme-to-heme ET (2.1); the classical molecular mechanics and molecular dynamics methods used to calculate the redox potentials and hence driving forces, as well as reorganization free energies - both needed in the Marcus rate expression for ET (2.2); methods employed to obtain electronic coupling matrix elements (also required in Marcus theory), namely density functional theory in general as well as in the context of coupling calculations (2.3); and empirical docking calculations to determine binding sites and affinities of a soluble ligand to a protein

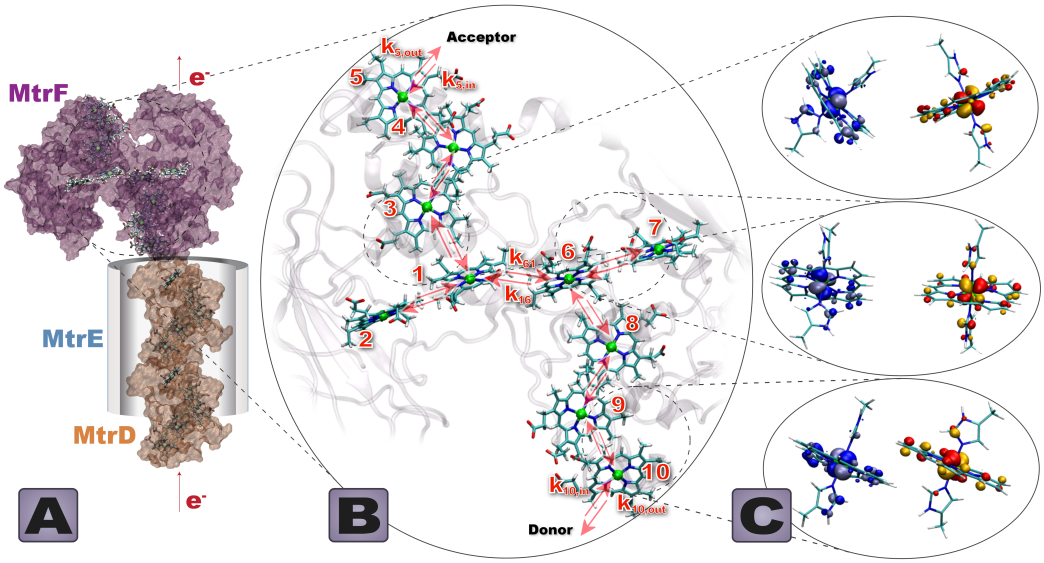

Figure 1.6: (A) A model of the transmembrane cytochrome complex MtrFDE. The OM deca-heme cytochrome MtrF (pdb code 3PMQ) is connected to the periplasmic/membrane cytochrome MtrD (modelled here by two NrfB proteins, pdb code 2OZY) via a porin MtrE (represented by a cartoon cylinder) enabling close contact between the two cytochromes across the outer membrane. (B) Arrangement of hemes in MtrF. Arrows denote single ET steps from heme i to heme j with rate constant k_{ji} (denoted exemplarily for pair 1-6) as well as ET steps to/from an external electron acceptor/donor, $k_{i,\mathrm{out}}$ and $k_{i,\mathrm{in}}$, shown exemplarily for hemes $i = 10$ and 5. (C) The three different heme pair motifs found in MtrF. From top to bottom: T-shaped, coplanar, stacked. d_π orbitals involved in the coupling are also depicted (see Chapter 3). Figure reprinted from [4]. (Figure created in collaboration with Cortland Johnson and Julian Breuer.)

(2.4).

Chapter 3, the main chapter of this thesis, presents our work on heme-to-heme ET in MtrF: Free energies, i. e. redox potentials for individual hemes and reorganization free energies for ET between adjacent hemes (3.1); and electronic couplings between adjacent hemes with the resulting heme-to-heme ET rates (3.2).

Chapter 4 presents the application of our parameters for heme-to-heme ET to through-protein transport. We model both an experimental setup under physiological conditions featuring cytochrome complexes spanning a lipid membrane (4.1); and recent scanning tunneling microscopy experiments measuring the current I through deca-heme cytochromes as a function of the external voltage V (4.2).

Chapter 5 turns to the interaction with soluble substrates, namely the interaction of MtrC with the proposed electron shuttle flavin mononucleotide (FMN). We validate and employ an empirical docking protocol to investigate FMN docking to MtrC; and carry out Simulated Annealing studies on MtrC

to discover potential conformational changes upon reduction and cleavage of a disulphide bond which in experiment has been reported to dramatically increase the affinity for FMN.

Finally, we summarize our findings in Chapter 6 and provide an outlook on extensions and continuations of the work presented here.

Chapter 2

Theory

2.1 Semi-classical Marcus theory

Electron transfer (ET) over distances like those typical for heme groups in a protein is well described by semi-classical nonadiabatic Marcus theory [50, 51]. This theory states the ET reaction rate constant k_{ET} between two redox partners as:

$$k_{ET} = \frac{1}{\hbar}\left\langle |H_{AB}|^2 \right\rangle \left(\frac{\pi}{\lambda k_B T}\right)^{\frac{1}{2}} \exp\left(-\frac{\left(\Delta A + \lambda\right)^2}{4\lambda k_B T}\right) \tag{2.1}$$

Here, ΔA is the reaction free energy, or driving force; λ is the so-called reorganization free energy describing the free energy necessary to distort the initial system to the final state (after ET) without actually transferring the electron (see further discussion below); and the "coupling matrix element" H_{AB} is a Hamiltonian element between two diabatic states A and B that describe the initial (pre-ET) and final (post-ET) state of the system, respectively:

$$H_{AB} = \left\langle \psi_A \left| \hat{H} \right| \psi_B \right\rangle \tag{2.2}$$

where $\hat{H}$ denotes the electronic Hamiltonian and ψ_M denotes the electronic wave function for diabatic state M, i. e. a state that does not remain an eigenstate of the electronic Hamiltonian as nuclear coordinates change (as is the case for adiabatic states) but rather maintains certain physical characteristics [52].

In particular, the diabatic states used to model electron transfer are defined as states that maintain the same charge on donor and acceptor even as nuclear coordinates are distorted to non-equilibrium configurations. As the electron jump from donor to acceptor is fast compared to the nuclear motions of both electron transfer partners and solvent [53], it has to happen in a nuclear configuration for which the vertical energy difference between initial and final

state, i. e. $\Delta E\left(\vec{R}\right) = E_B\left(\vec{R}\right) - E_A\left(\vec{R}\right)$, is zero. ($E_A\left(\vec{R}\right)$ and $E_B\left(\vec{R}\right)$ refer to the energy in the initial and final state, respectively, for a given nuclear configuration $\vec{R}$. "Vertical" thus refers to a process with fixed nuclear coordinates.) The vertical energy gap $\Delta E\left(\vec{R}\right)$ is determined by the polarization state of the redox partners' environment (solvent and, in this case, protein), the so-called "outer sphere" contribution, and by the geometry of donor and acceptor themselves, the "inner sphere" contribution. It thus allows to sum up the role of an ensemble of nuclei into one parameter and constitutes a useful reaction coordinate for the overall electron transfer process [54, 5]. The two diabatic states can be described by free energy curves along this reaction coordinate, with each point on a curve referring to an ensemble of configurations featuring the same value for $\Delta E\left(\vec{R}\right)$, with an associated free energy of [5]:

$$A_M(\Delta E) = -k_B T \ln p_M(\Delta E) + const \tag{2.3}$$

where M is A for the initial or B for the final state, respectively; k_B and T are Boltzmann's constant and absolute temperature, respectively; and $p_M(\Delta E)$ is the probability of the energy gap value ΔE in state M. These free energy curves are illustrated in Fig. 2.1.

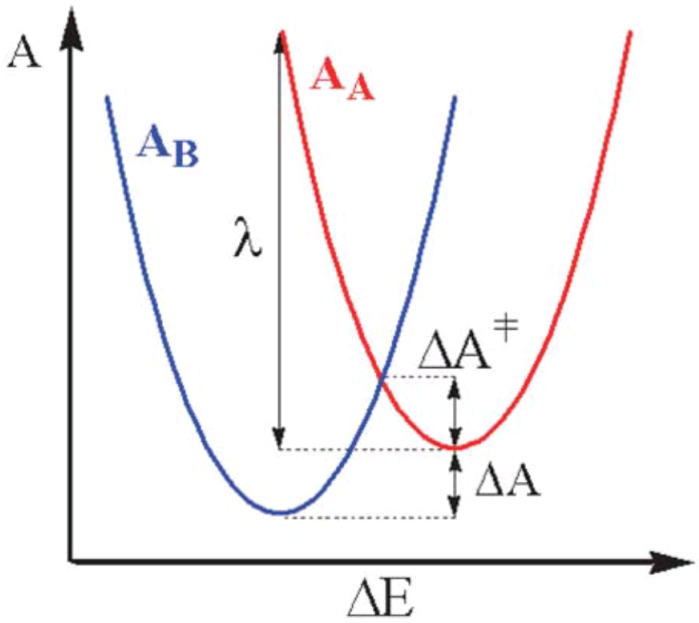

Figure 2.1: Free energy curves of diabatic initial and final state (A_A and A_B, respectively) of an electron transfer reaction along the reaction coordinate ΔE (vertical energy gap between both states). ΔA denotes the reaction free energy, $\Delta A^{\ddagger}$ the activation free energy and λ the reorganization free energy (see text). Reproduced from Ref. [5] with permission from the PCCP Owner Societies (figure prepared by Dr. Jochen Blumberger).

The use of ΔE as a reaction coordinate leads to a linear free energy relationship [5]:

$$A_B(\Delta E) - A_A(\Delta E) = \Delta E \tag{2.4}$$

I. e., the difference in free energy between initial and final state for each value of ΔE is equal to ΔE itself. Nonadiabatic electron transfer in the Marcus picture can then be described as follows: In the initial state (red curve in figure 2.1), solvent polarization and cofactor geometries correspond to the reduced donor and oxidized acceptor, with ΔE fluctuating around the equilibrium value for the initial state. (Rare) thermal fluctuations enable the system to reach the intersection point of the two free energy curves where the free energies of initial and final state are equal, ΔE vanishes (by virtue of Eq. 2.4) and electron hopping becomes possible. To reach such a configuration however requires the activation free energy $\Delta A^{\ddagger}$; this energy in turn is related to the reorganization free energy λ, the energy that would be required to distort the system in the initial state to the equilibrium ensemble of geometries of the final state, but without actually transferring the electron. (This is depicted in Fig. 2.1 by moving along the red curve to the left until reaching the point above the minimum of the blue curve.)

If the probability distribution of energy gaps p_A is Gaussian, the corresponding free energy curve A_A becomes harmonic, as does A_B due to Eq. 2.4 (with both parabolas having the same curvature). As a consequence, the reaction free energy ΔA can be expressed as the average of the ensemble averages of ΔE in the initial and final state [5]:

$$\Delta A = \frac{(\langle \Delta E \rangle_A + \langle \Delta E \rangle_B)}{2} \tag{2.5}$$

where $\langle \Delta E \rangle_M$ denotes the ensemble average of ΔE in state M. In addition, the reorganization energy λ can be expressed similarly as [5]:

$$\lambda = \frac{(\langle \Delta E \rangle_A - \langle \Delta E \rangle_B)}{2} \tag{2.6}$$

The activation free energy $\Delta A^{\ddagger}$ can then be calculated from the vertex of the initial state's parabola and the intersection of the two parabolas; this yields:

$$\Delta A^{\ddagger} = \frac{\left(\Delta A + \lambda\right)^2}{4\lambda} \tag{2.7}$$

Once the system does reach the transition state ensemble via thermal fluctuations, it may or may not cross to the final state. The probability of the transition happening, i. e. ET taking place, is determined by the aforementioned electronic coupling matrix element H_{AB} - more precisely, by the thermal average of its square, $\left\langle |H_{AB}|^2 \right\rangle$. H_{AB} is averaged in the ensemble at the

transition point, i. e. in the ensemble with $\Delta E\left(\vec{R}\right) = 0$. (For these are the nuclear configurations where ET can take place only.)

The electronic coupling term H_{AB} has been found to decay exponentially with distance [16]. λ usually varies between 0.25 and 1.5 eV and for most proteins with solvent-excluded cofactors assumes values in the range of 0.6 to 0.9 eV [55]; no large variations in the driving force are usually observed as efficient energy transfer requires small driving forces [5].

Experimentally, the free energy of electron transfer can often be determined from the redox potential difference of the involved cofactors. Although λ cannot be simply determined from the temperature dependence of the rate constant in a reliable way [56], the driving force can be varied (for constant temperatures) by modifying the cofactors' chemical substituents; this, however, rests on the assumption that $\left\langle |H_{AB}|^2 \right\rangle$ remains unaffected by the change in the chemical environment of the cofactors [5]. This makes computational approaches particularly appealing as these provide an alternative way of estimating the quantities needed to describe electron transfer reactions by Marcus theory. They also have the advantage of enabling decomposition of these quantities into individual contributions from different components (cofactors, individual residues, solvent) [5].

The next two subsections therefore discuss how reaction and reorganization free energies as well as coupling elements can be obtained via computational methods. A purely classical description of the system was used in the calculation of ΔA and λ; however, the calculation of H_{AB} requires quantum chemical methods. Still, $|H_{AB}|^2$ needs to be averaged over an ensemble of nuclear configurations, and these configurations could be sampled classically as well. Section 2.2 therefore describes the methods used in obtaining ΔA and λ as well as in sampling nuclear configurations for H_{AB}. The actual calculation of H_{AB} for single configurations using quantum chemical methods is covered in Section 2.3.

2.2 Classical methods

2.2.1 Sampling the phase space - Molecular Dynamics

All three quantities needed to establish the heme-to-heme ET rates k_{ET} - reaction free energies ΔA, reorganization free energies λ and electronic coupling elements $\left\langle |H_{AB}|^2 \right\rangle$ require the evaluation of some ensemble average. Two decisions have to be made in this regard:

1. The computational model used to describe the system and its Hamiltonian $\mathscr{H}$;

2. The method used to sample the system's configuration space in order to obtain ensemble averages.

In regard to the first point, the configurations to be sampled belong to the equilibrium ensemble of a protein system without any chemical changes, i. e. formation or breaking of chemical bonds. Therefore, costly electronic structure methods can be avoided in favor of the computationally much cheaper Molecular Mechanics methods [57]. In constrast to electronic structure methods, the approach of Molecular Mechanics does not deal with the electronic degrees of freedom explicitly; the Hamiltonian of a molecular system is expressed as a function of nuclear coordinates and momenta, possibly augmented by additional terms describing electronic degrees of freedom in a very simplified manner, like induced point dipoles [58]. In the absence of such additional terms (referred to as the "nonpolarizable" case in the following due to the absence of electronic polarization), the Hamiltonian is generally given by:

$$\mathscr{H}\left(\vec{R},\vec{P}\right) = K\left(\vec{P}\right) + V\left(\vec{R}\right) = \sum_{i=1}^{N} \frac{|\vec{p}_i|^2}{2m_i} + V\left(\vec{R}\right) \tag{2.8}$$

i. e. it can be expressed as a sum of kinetic energy $K\left(\vec{P}\right)$ and potential energy $V\left(\vec{R}\right)$, with $\vec{R}$ and $\vec{P}$ denoting the total of all position and momentum vectors, respectively, and $\vec{p}_i$ and m_i momentum and mass of atom i, respectively; the sum runs over all N atoms in the system.

Different functional forms for the potential energy $V\left(\vec{R}\right)$ in this purely classical model Hamiltonian exist; they all contain parameters that need to be determined by fitting to either experimental observables (e.g. vibrational frequencies) or quantum mechanical calculations (on simple model compounds) [59]. Functional form and parameters together then constitute a so-called force field.

A common functional form for a nonpolarizable molecular mechanical potential energy (and indeed the form used for the sampling of configuration space in this work) is given by [60]:

$$\begin{aligned} V\left(\vec{R}\right) &= \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angles} \frac{\kappa_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsions} \sum_{n}^{M} \frac{V_n}{2}\left(1 + \cos\left(n\omega - \gamma\right)\right) \\ &+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right) + \frac{q_i \cdot q_j}{4\pi\varepsilon_0 r_{ij}} \right) \end{aligned} \tag{2.9}$$

In this equation, the total potential energy of the molecular system is decomposed into bonding terms (first three summations) and nonbonding terms (last double-summation). The bonding terms assume that each covalent bond

length l_i and angle θ_i between covalent bonds fluctuates around an equilibrium value ($l_{i,0}$ for bonds and $\theta_{i,0}$ for angles, respectively) and that for values close to the respective equilibrium value (for which the force field is designed only), the associated potential energy can be approximated to be harmonic with a force constant k_i (for bonds) or κ_i (for angles), respectively. Torsional terms (third summation) represent interactions between atoms separated by three bonds in dependence on rotation around the second (central) bond; they thus constitute four-body interaction terms of these atoms and the two atoms connecting them and are used to model steric and, to some extent, nonbonded (see below) interactions between the two outer atoms. Torsional terms are modelled as a sum of M cosine terms (with M often just equal to 1) of different multiplicity n, "barrier height" V_n and phase shift γ; ω denotes the torsional angle, i.e. the angle of rotation around the bond between the two inner atoms.

The last double-summation runs over all unique pairs of atoms in the system that are at least three bonds apart from each other (or not connected by bonds at all, respectively); this sum represents electric Coulomb interactions (second addend) as well as Van der Waals interactions (first addend), the latter being interactions that are even present in completely noncharged and unpolar systems. Coulomb interactions are calculated by assigning effective (partial) atomic charges q_i to each atom and calculating the interaction using Coulomb's law, with r_{ij} as the interatomic distance and ε_0 as the electric constant; Van der Waals interactions are modelled by the so-called Lennard-Jones potential which consists of an attractive long-range dispersion interaction term $-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6$ as well as a short-range repulsion term $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}$, both together resulting in a potential well of depth ε_{ij}; σ_{ij} is the interatomic distance for which the two parts cancel out. As mentioned above, torsional interaction terms already partially account for nonbonded interactions as well; thus Van der Waals and Coulomb interactions are typically scaled by a scaling factor smaller than one for atoms separated by three bonds. As bond and angle terms are supposed to fully describe the interactions between the atoms involved, nonbonded interactions are not calculated at all for atoms separated by only one or two bonds.

In order to avoid artificial boundary effects at the system's boundary, periodic boundary conditions [61] are frequently used and were used in this work as well. In this approach, the simulation box is surrounded by an infinite set of identical copies ("image cells"), filling out the space entirely. If an atom leaves the central cell, its image enters on the opposite side (i. e. the atom is projected back into the central cell again). Electrostatic interactions are then calculated not only within the central cell but also between atoms in the

central cell and all image cells:

$$V_{elec}\left(\vec{R}\right) = \frac{1}{2}\sum_{\vec{I}}^{\prime}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{q_i \cdot q_j}{4\pi\varepsilon_0\left|\vec{r_{ij}}+\vec{I}\right|}\right) \tag{2.10}$$

where $V_{elec}\left(\vec{R}\right)$ denotes the electrostatic potential energy; the sum $\sum_{\vec{I}}^{\prime}$ runs over all cells, i.e. central cell as well as image cells, where $\vec{I}$ is the translation vector of a given image cell ($\vec{0}$ for the central cell itself); the interatomic distance r_{ij} is substituted by $\left|\vec{r_{ij}}+\vec{I}\right|$, the distance between an atom i in the central cell and atom j's image in the image cell specified by $\vec{I}$. The apostrophe above the summation sign indicates that the interaction of an atom with itself in the central cell is still ignored, but atoms do interact with their own periodic images. To calculate this infinite sum, Ewald summation is frequently used [62, 63]. Van der Waals interactions are also computed between atoms in the central cell and surrounding neighbour images; however, a cutoff is usually applied for Van der Waals interactions to prevent an atom from interacting with any other atom twice [64].

This description of the electrostatic energy however disregards any polarization effects on individual atoms from their environment. Several simple procedures exist to incorporate such effects at least partially, like fluctuating atomic charges and the aforementioned induced dipoles [58]. In the latter approach, each atom is assigned an isotropic atomic polarizability α_i and a polarization energy term E_{pol} is added to the total energy, defined as [58]

$$E_{pol} = -\frac{1}{2}\sum_{i}\alpha_i\vec{E}_i^{(0)}\vec{E}_i \tag{2.11}$$

where $\vec{E}_i$ denotes the total electric field acting on atom i (from both static monopole charges and induced dipoles) and $\vec{E}_i^{(0)}$ denotes the field due to monopole charges only.

Coming to the choice on how to sample the phase space, there are two particularly popular computational approaches: Molecular Dynamics [65] and Monte Carlo methods [66]. Monte Carlo methods are based on randomly generating new configurations of atoms in a system and accepting or rejecting these new configurations with a probability depending on their energy, thereby obtaining a Boltzmann (or any other desired) distribution. However, for molecular systems most random coordinate changes will sharply increase the system's energy by e.g. severely stretching bonds and thus will lead to configurations with small Boltzmann weights; this renders the application of Monte Carlo methods to molecular systems rather difficult [67]. Molecular Dynamics, on the other side, are based on computing actual dynamics of a

system in order to obtain a trajectory in phase space along which the quantities of interest can be computed; the ergodic hypothesis can then be used to equate a quantity's time average as obtained from these dynamics with its ensemble average. (Provided that the average from these dynamics of finite length can be assumed to come sufficiently close to the average that infinitely long dynamics would yield.) The modelling of a system via Molecular Mechanics allows to describe the system's dynamics by the classical Newtonian dynamics of its atoms; the equation to be integrated is then Newton's equation of motion for each atom:

$$\frac{d^2\vec{r}_i}{dt^2} = \frac{\vec{F}_i}{m_i} \tag{2.12}$$

where $\vec{r}_i$ denotes the position vector of atom i, m_i denotes its mass, $\frac{d^2}{dt^2}$ denotes the second derivative with respect to time and $\vec{F}_i$ is the external force acting on atom i; this force is in turn obtained from the potential energy's gradient with respect to atom i's coordinates:

$$\vec{F}_i = -\vec{\nabla}_{\vec{r}_i} V\left(\vec{R}\right) \tag{2.13}$$

As for a large system with a molecular mechanical force field solving these equations analytically becomes unfeasible, numerical finite difference methods have to be employed which are based on discretizing time into tiny steps. The algorithm used in this work is the so-called velocity Verlet algorithm [68]; in this algorithm, positions are updated based on positions $\vec{r}_i$, velocities $\vec{v}_i$ and accelerations $\vec{a}_i$ at the current time t, whereas velocities can be updated once the forces and hence accelerations at the next step $t + \delta t$ are known:

$$\vec{r}_i(t+\delta t) = \vec{r}_i(t) + \delta t \cdot \vec{v}_i(t) + \frac{1}{2}\delta t^2 \cdot \vec{a}_i(t) \tag{2.14}$$

$$\vec{v}_i(t+\delta t) = \vec{v}_i(t) + \frac{1}{2}\delta t \cdot (\vec{a}_i(t) + \vec{a}_i(t+\delta t)) \tag{2.15}$$

where δt denotes the discrete time step of integration.

Incorporation of polarizability via induced point dipoles increases the computational cost of force field calculations considerably; in this work, only a nonpolarizable force field was used in sampling the phase space.

2.2.2 Redox potentials

The reaction free energies (or driving forces) for the individual electron transfer steps in MtrF were calculated indirectly via the redox potentials of the in-

dividual cofactors. The redox potential of a heme cofactor is the difference in free energy between the state of a (protein) system with that cofactor oxidized and reduced; this free energy difference can be computed via the method of Thermodynamic Integration (TI) . The difference of the redox potentials of two cofactors then yields the driving force for ET between them. (The term "redox potential" is used in this work as a synonym for the free energy of removing an electron from a heme cofactor; it is thus equal to the actual redox potential (in Volt) multiplied by the charge of one electron. However, as the electronvolt is used as the unit of energy throughout this work, the numerical values are identical.) In principle, the redox potential of one heme in MtrF depends on the redox states of the other hemes. We chose to treat all other hemes as oxidized so that our redox potentials would apply to the transport of one excess electron through the protein.

In the following, we outline the TI formalism. The free energy of a system being a state function, the free energy difference ΔG between the two redox states of interest can always be written as an integral over an arbitrary "coupling parameter" ε:

$$\Delta G = G_1 - G_0 = \int_0^1 \frac{dG_\varepsilon}{d\varepsilon} d\varepsilon \tag{2.16}$$

where G_0 denotes the Gibbs free energy of the initial (reduced) state, G_1 denotes the Gibbs free energy of the final (oxidized) state, $\frac{dG_\varepsilon}{d\varepsilon}$ is the derivative of the free energy which is taken to be a function of ε and the integral limits denote an integral from initial to final state (i.e. $\varepsilon = 0$ corresponds to the initial, $\varepsilon = 1$ to the final state). The Gibbs free energy is used rather than the Helmholtz free energy since the system to be modelled is a protein located on the outside of a biological cell, i.e. in contact with an aqueous solution and thus under isobaric conditions. For the general relation 2.16 to have a practial utility, an expression for the derivative $\frac{dG_\varepsilon}{d\varepsilon}$ needs to be found. To this end, a hypothetical intermediate state with a free energy G_ε and a Hamiltonian $\mathscr{H}_\varepsilon$ is defined where $\mathscr{H}_\varepsilon$ is a function of the coupling parameter ε and G_ε depends on ε via $\mathscr{H}_\varepsilon$:

$$G_\varepsilon = -\frac{1}{\beta} \ln Q_\varepsilon \tag{2.17}$$

$$Q_\varepsilon = \frac{1}{N! h^{3N} V_0} \int \exp\left(-\beta \left(\mathscr{H}_\varepsilon\left(\vec{R}, \vec{P}, \varepsilon\right) + pV\right)\right) d\vec{R} d\vec{P} dV \tag{2.18}$$

where N denotes the number of particles in the system, h denotes Planck's constant, $\beta = \frac{1}{k_B T}$ with k_B as Boltzmann's constant and T as the system tem-

perature, p and V denote pressure and volume, respectively, V_0 is some basic unit of volume (whose exact choice is not important for the following considerations), $\vec{R}$ serves as a short form for all individual particle position vectors and $\vec{P}$ serves as a short form for all individual particle momentum vectors [69]. The following procedure has been described in Ref. [70] using the Helmholtz free energy and is adapted here to use the Gibbs free energy. Taking the derivative of G_ε with respect to ε yields:

$$\frac{dG_\varepsilon}{d\varepsilon} = -\frac{1}{\beta}\frac{\frac{\partial Q_\varepsilon}{\partial \varepsilon}}{Q_\varepsilon} \tag{2.19}$$

As Q_ε only depends on ε via $\mathscr{H}_\varepsilon$, $\frac{\partial Q_\varepsilon}{\partial \varepsilon}$ can be expressed as:

$$\begin{aligned}\frac{\partial Q_\varepsilon}{\partial \varepsilon} &= \frac{1}{N!h^{3N}V_0}\int \frac{\partial}{\partial \varepsilon}\exp\left(-\beta\left(\mathscr{H}_\varepsilon\left(\vec{R},\vec{P},\varepsilon\right)+pV\right)\right)d\vec{R}d\vec{P}dV \\ &= -\frac{\beta}{N!h^{3N}V_0}\int \frac{\partial \mathscr{H}_\varepsilon\left(\vec{R},\vec{P},\varepsilon\right)}{\partial \varepsilon}\exp\left(-\beta\left(\mathscr{H}_\varepsilon\left(\vec{R},\vec{P},\varepsilon\right)+pV\right)\right)d\vec{R}d\vec{P}dV\end{aligned} \tag{2.20}$$

and thus:

$$\begin{aligned}\frac{dG_\varepsilon}{d\varepsilon} &= \frac{\frac{1}{N!h^{3N}V_0}\int \frac{\partial \mathscr{H}_\varepsilon\left(\vec{R},\vec{P},\varepsilon\right)}{\partial \varepsilon}\exp\left(-\beta\left(\mathscr{H}_\varepsilon\left(\vec{R},\vec{P},\varepsilon\right)+pV\right)\right)d\vec{R}d\vec{P}dV}{Q_\varepsilon} \\ &= \left\langle \frac{\partial \mathscr{H}_\varepsilon}{\partial \varepsilon}\right\rangle_\varepsilon\end{aligned} \tag{2.21}$$

i.e. the derivative of the free energy with respect to ε for a certain value of ε can be computed as the average of $\frac{\partial \mathscr{H}_\varepsilon}{\partial \varepsilon}$ in the isothermal-isobaric ensemble specified by the given value of ε. As only the potential energy part of the Hamiltonian depends on ε, the change in free energy between the initial and final state is then given by:

$$\Delta G = \int_0^1 \left\langle \frac{\partial V_\varepsilon}{\partial \varepsilon}\right\rangle_\varepsilon d\varepsilon \tag{2.22}$$

where V_ε denotes the potential energy as a function of ε, and the ensemble average is in fact an average over configuration space (as the momenta can be integrated out). ΔG can then be calculated by defining V_ε and its dependency on ε and choosing a set of integration points, or windows, in the ε range of 0 to 1 for which to obtain the ensemble average $\left\langle \frac{\partial V_\varepsilon}{\partial \varepsilon}\right\rangle_\varepsilon$ in order to evaluate

Eq. 2.22 numerically.

There are in principle several possibilities of defining V_ε and its dependency on ε. $V_\varepsilon\left(\vec{R}\right)$ could be defined to be a weighted average of initial and final potential energy function $V_0\left(\vec{R}\right)$ and $V_1\left(\vec{R}\right)$, respectively, without using an explicit expression for $V_\varepsilon\left(\vec{R}\right)$:

$$V_\varepsilon\left(\vec{R}\right) := (1-\varepsilon)\, V_0\left(\vec{R}\right) + \varepsilon \cdot V_1\left(\vec{R}\right) \tag{2.23}$$

In this approach, both potential functions $V_0\left(\vec{R}\right)$ and $V_1\left(\vec{R}\right)$ are needed during the simulation. This would result in a particularly simple expression for $\frac{\partial V_\varepsilon}{\partial \varepsilon}$:

$$\frac{\partial V_\varepsilon\left(\vec{R}\right)}{\partial \varepsilon} = \frac{\partial}{\partial \varepsilon}\left((1-\varepsilon)\, V_0\left(\vec{R}\right) + \varepsilon \cdot V_1\left(\vec{R}\right)\right) = V_1\left(\vec{R}\right) - V_0\left(\vec{R}\right) \tag{2.24}$$

However, this approach was not possible in the present work as the software used (see Chapter 3) did not support the simultaneous use of two potential energy functions. Thus, $V_\varepsilon\left(\vec{R}\right)$ was instead defined by making the force field parameters themselves dependent on the coupling parameter ε (with $\varepsilon = 0$ and $\varepsilon = 1$ reproducing the parameters for the initial and final state, respectively). In the force field parameters used to model the heme cofactors [71] in this work, the atomic charges are the only parameters different between the two oxidation states. Thus, it is convenient to define the intermediate states for Thermodynamic Integration by linearly scaling (or "morphing") the charges on a heme cofactor:

$$q_{i,\varepsilon} = (1-\varepsilon)\, q_{i,0} + \varepsilon \cdot q_{i,1} \tag{2.25}$$

where $q_{i,0}$ and $q_{i,1}$ are the atomic charges of atom i in initial (reduced) and final (oxidized) state, respectively, and $q_{i,\varepsilon}$ denotes the morphed charge in the intermediate state given by ε. Based on this linear charge morphing, an expression for $\frac{\partial V_\varepsilon}{\partial \varepsilon}$ has been derived by the author of this work. This derivation is presented in the following.

With atomic charges being the only set of parameters depending on the scaling parameter ε, the derivative $\frac{\partial V_\varepsilon}{\partial \varepsilon}$ correspondingly only contains electrostatic contributions:

$$\frac{\partial V_\varepsilon}{\partial \varepsilon} = \frac{\partial V_{\varepsilon,elec}}{\partial \varepsilon} \tag{2.26}$$

where $V_{\varepsilon,elec}\left(\vec{R}\right)$ denotes the total electrostatic energy for an intermediate state and is given by:

$$
\begin{aligned}
V_{\varepsilon,elec}\left(\vec{R}\right) &= \frac{1}{2}\sum_{\vec{I}}^{\prime}\sum_{i=1}^{K}\sum_{k=1}^{K}\left(\frac{q_{i,\varepsilon}\cdot q_{k,\varepsilon}}{4\pi\varepsilon_0\left|\vec{r_{ik}}+\vec{I}\right|}\right) \\
&+ \sum_{\vec{I}}^{\prime}\sum_{i=1}^{K}\sum_{m=1}^{M}\left(\frac{q_{i,\varepsilon}\cdot q_m}{4\pi\varepsilon_0\left|\vec{r_{im}}+\vec{I}\right|}\right)+\frac{1}{2}\sum_{\vec{I}}^{\prime}\sum_{m=1}^{M}\sum_{n=1}^{M}\left(\frac{q_m\cdot q_n}{4\pi\varepsilon_0\left|\vec{r_{mn}}+\vec{I}\right|}\right)
\end{aligned}
\tag{2.27}
$$

where atom indices i and k denote atoms whose charges change (total number of these atoms: K) and indices m and n denote atoms whose charge is constant (total number: M). The first sum includes all interactions between atoms with changing charges, the second term includes all interactions between atoms whose charge changes and those whose charge is constant (disjoint sets of atoms, thus no double-counting in this case) and the last sum describes the interactions between all atoms with constant charges. In the first and third sum, interactions of an atom with itself in the central cell are again excluded. The derivative $\frac{\partial V_{\varepsilon,elec}}{\partial\varepsilon}$ then reduces to the derivative of the first two sums with respect to ε as the coupling parameter does not appear in the last term. Furthermore, ε only appears in the numerator (of each individual pair interaction) in the first two sums. It is thus sufficient to evaluate the derivative for the numerators only. The second numerator gives:

$$
\begin{aligned}
q_{i,\varepsilon}\cdot q_m &= \left((1-\varepsilon)\,q_{i,0}+\varepsilon\cdot q_{i,1}\right)q_m \\
&= \varepsilon\left(q_{i,1}\cdot q_m - q_{i,0}\cdot q_m\right)+q_{i,0}\cdot q_m \qquad (2.28)\\
\frac{\partial\left(q_{i,\varepsilon}\cdot q_m\right)}{\partial\varepsilon} &= q_{i,1}\cdot q_m - q_{i,0}\cdot q_m \qquad (2.29)
\end{aligned}
$$

The first numerator requires a few more steps. Expanding the product and bracketing out same powers of ε yields:

$$
\begin{aligned}
q_{i,\varepsilon}\cdot q_{k,\varepsilon} &= \left(\varepsilon\cdot q_{i,1}-\varepsilon\cdot q_{i,0}+q_{i,0}\right)\cdot\left(\varepsilon\cdot q_{k,1}-\varepsilon\cdot q_{k,0}+q_{k,0}\right) \\
&= \varepsilon^2\left(q_{i,1}\cdot q_{k,1}-q_{i,1}\cdot q_{k,0}-q_{i,0}\cdot q_{k,1}+q_{i,0}\cdot q_{k,0}\right) \\
&+ \varepsilon\left(q_{i,1}\cdot q_{k,0}+q_{i,0}\cdot q_{k,1}-2q_{i,0}\cdot q_{k,0}\right) \\
&+ q_{i,0}\cdot q_{k,0} \\
&+ \varepsilon q_{i,1}\cdot q_{k,1}-\varepsilon q_{i,1}\cdot q_{k,1} \qquad (2.30)
\end{aligned}
$$

where at the end 0 has been added to further transform the right hand side (RHS). Exchanging $q_{i,0}\cdot q_{k,0}$ for $q_{i,1}\cdot q_{k,1}$ in the second bracket on the RHS

and factorizing yields:

$$\begin{aligned}
q_{i,\varepsilon}\cdot q_{k,\varepsilon} &= \varepsilon^2\left(q_{i,1}\cdot q_{k,1}-q_{i,1}\cdot q_{k,0}-q_{i,0}\cdot q_{k,1}+q_{i,0}\cdot q_{k,0}\right)\\
&+ \varepsilon\left(q_{i,1}\cdot q_{k,0}+q_{i,0}\cdot q_{k,1}-q_{i,0}\cdot q_{k,0}-q_{i,1}\cdot q_{k,1}\right)\\
&+ q_{i,0}\cdot q_{k,0}+\varepsilon q_{i,1}\cdot q_{k,1}-\varepsilon q_{i,0}\cdot q_{k,0}\\
&= \varepsilon^2\left(q_{i,1}-q_{i,0}\right)\left(q_{k,1}-q_{k,0}\right)\\
&- \varepsilon\left(q_{i,1}-q_{i,0}\right)\left(q_{k,1}-q_{k,0}\right)\\
&+ \varepsilon\left(q_{i,1}\cdot q_{k,1}-q_{i,0}\cdot q_{k,0}\right)\\
&+ q_{i,0}\cdot q_{k,0}\\
&= \left(\varepsilon^2-\varepsilon\right)\Delta q_i\Delta q_k\\
&+ \varepsilon\left(q_{i,1}\cdot q_{k,1}-q_{i,0}\cdot q_{k,0}\right)\\
&+ q_{i,0}\cdot q_{k,0} \qquad (2.31)
\end{aligned}$$

where

$$\Delta q_i := \left(q_{i,1}-q_{i,0}\right) \qquad (2.32)$$

$$\Delta q_k := \left(q_{k,1}-q_{k,0}\right) \qquad (2.33)$$

Taking the derivative now yields:

$$\frac{\partial\left(q_{i,\varepsilon}\cdot q_{k,\varepsilon}\right)}{\partial\varepsilon} = \left(2\varepsilon-1\right)\Delta q_i\Delta q_k+q_{i,1}\cdot q_{k,1}-q_{i,0}\cdot q_{k,0} \qquad (2.34)$$

Having derivatives for the numerators of the first two sums in 2.27, the combined derivative can now be obtained by substituting the numerators in the RHS of 2.27 by these derivatives:

$$\begin{aligned}
\frac{\partial V_{\varepsilon,elec}\left(\vec{R}\right)}{\partial\varepsilon} &= \frac{1}{2}\sum_{\vec{I}}^{\prime}\sum_{i=1}^{K}\sum_{k=1}^{K}\left(\frac{\left(2\varepsilon-1\right)\Delta q_i\Delta q_k+q_{i,1}\cdot q_{k,1}-q_{i,0}\cdot q_{k,0}}{4\pi\varepsilon_0\left|\vec{r_{ik}}+\vec{I}\right|}\right)\\
&+ \sum_{\vec{I}}^{\prime}\sum_{i=1}^{K}\sum_{m=1}^{M}\left(\frac{q_{i,1}\cdot q_m-q_{i,0}\cdot q_m}{4\pi\varepsilon_0\left|\vec{r_{im}}+\vec{I}\right|}\right)\\
&= V_{1,elec}\left(\vec{R}\right)-V_{0,elec}\left(\vec{R}\right)+\left(2\varepsilon-1\right)V_{\Delta,elec}\left(\vec{R}\right) \qquad (2.35)
\end{aligned}$$

with:

$$
\begin{aligned}
V_{1,elec}\left(\vec{R}\right) &= \frac{1}{2}\sum_{\vec{I}}{}' \sum_{i=1}^{K}\sum_{k=1}^{K}\left(\frac{q_{i,1}\cdot q_{k,1}}{4\pi\varepsilon_0\left|\vec{r_{ik}}+\vec{I}\right|}\right) + \sum_{\vec{I}}{}' \sum_{i=1}^{K}\sum_{m=1}^{M}\left(\frac{q_{i,1}\cdot q_m}{4\pi\varepsilon_0\left|\vec{r_{im}}+\vec{I}\right|}\right) \\
&+ \frac{1}{2}\sum_{\vec{I}}{}' \sum_{m=1}^{M}\sum_{n=1}^{M}\left(\frac{q_m\cdot q_n}{4\pi\varepsilon_0\left|\vec{r_{mn}}+\vec{I}\right|}\right) \\
V_{0,elec}\left(\vec{R}\right) &= \frac{1}{2}\sum_{\vec{I}}{}' \sum_{i=1}^{K}\sum_{k=1}^{K}\left(\frac{q_{i,0}\cdot q_{k,0}}{4\pi\varepsilon_0\left|\vec{r_{ik}}+\vec{I}\right|}\right) + \sum_{\vec{I}}{}' \sum_{i=1}^{K}\sum_{m=1}^{M}\left(\frac{q_{i,0}\cdot q_m}{4\pi\varepsilon_0\left|\vec{r_{im}}+\vec{I}\right|}\right) \\
&+ \frac{1}{2}\sum_{\vec{I}}{}' \sum_{m=1}^{M}\sum_{n=1}^{M}\left(\frac{q_m\cdot q_n}{4\pi\varepsilon_0\left|\vec{r_{mn}}+\vec{I}\right|}\right) \\
V_{\Delta,elec}\left(\vec{R}\right) &:= \frac{1}{2}\sum_{\vec{I}}{}' \sum_{i=1}^{K}\sum_{k=1}^{K}\left(\frac{\Delta q_i \Delta q_k}{4\pi\varepsilon_0\left|\vec{r_{ik}}+\vec{I}\right|}\right)
\end{aligned}
$$

In this derivation, 0 has been added in between on the right hand side (by adding and subtracting the sum over m and n) in order to yield the expressions for $V_{1,elec}\left(\vec{R}\right)$ and $V_{0,elec}\left(\vec{R}\right)$, respectively. $V_{\Delta,elec}\left(\vec{R}\right)$ describes the electrostatic energy of a system with point charges only on atoms whose charge changes between the two states, the point charges' magnitude being given by the charge difference between initial and final state for each of these atoms.

Thus, the derivative 2.35 differs from Eq. 2.24 only in the additional term $(2\varepsilon-1)V_{\Delta,elec}\left(\vec{R}\right)$. (And by using the electrostatic rather than overall potential energies; however, as mentioned before, $\frac{\partial V_{\varepsilon,elec}(\vec{R})}{\partial\varepsilon} = \frac{\partial V_{\varepsilon}(\vec{R})}{\partial\varepsilon}$, and furthermore $V_1\left(\vec{R}\right) - V_0\left(\vec{R}\right) = V_{1,elec}\left(\vec{R}\right) - V_{0,elec}\left(\vec{R}\right)$.) This additional term can easily be computed for each snapshot by assigning the aforementioned charge differences for each atom to that atom's position and calculating the electrostatic energy for that artificial system the same way it is computed for the actual system.

2.2.3 Reorganization free energies

The reorganization free energy λ for a given ET process consists of contributions from two different regions:

$$\lambda = \lambda_{inner} + \lambda_{outer} \tag{2.36}$$

where λ_{inner} describes the contribution of the so-called inner sphere, i. e. donor and acceptor themselves; and λ_{outer} describes the contribution from

their environment, the so-called outer sphere. The latter consists of the change in polarization of the surrounding medium (in our case, protein and solvent) while the former describes the reorganization (nuclear and electronic) within the cofactors. Hence, while the outer sphere contribution can be well described with a classical force field, the inner sphere contribution requires quantum chemical calculations. Fortunately, previous work has shown [5] that these two components can as a very good approximation be treated as independent from each other; that is, the inner sphere contribution is largely specific to a cofactor and not dependent on its environment. Hence, rather than calculating the inner sphere contribution explicitly, we can apply a value $\lambda_{inner} = 0.05$ eV derived from previous work. [9]

The outer sphere contribution λ_{outer} is then calculated via Eq. 2.6 for which we need to sample the vertical energy gap $\Delta E\left(\vec{R}\right) = E_B\left(\vec{R}\right) - E_A\left(\vec{R}\right)$ in the initial state A and final state B of ET between two cofactors (whereas $\frac{\partial V_\varepsilon}{\partial \varepsilon}$ contained the energy gap $V_1\left(\vec{R}\right) - V_0\left(\vec{R}\right)$ for *oxidation* of a single cofactor). To avoid double-counting, we then need to subtract the *classical* inner sphere contribution from the classically obtained λ (Eq. 2.6), yielding a purely classical λ_{outer}. [9] This is done by subtracting from each energy gap ΔE the contribution of the cofactors only, yielding the outer sphere energy gap ΔE_{outer} which is then used for Eq. 2.6:

$$\Delta E_{outer} = E_B\left(\vec{R}\right) - E_A\left(\vec{R}\right) - \left(\Delta E^D_{inner} - \Delta E^A_{inner}\right) \tag{2.37}$$

$$\Delta E^C_{inner} = \sum_{i,k=1,i\neq k}^{K} \frac{q_{i,1}q_{k,1} - q_{i,0}q_{k,0}}{4\pi\varepsilon_0 \left|\vec{r_{ik}}\right|} \tag{2.38}$$

Here, A and B denote the initial and final state of ET, respectively (as in Section 2.1); C denotes either the donor (D) or acceptor (A); the indices 0 and 1 refer to the reduced (0) and oxidixed (1) state of a cofactor; and the indices i and k in the definition of ΔE^C_{inner} run over the inner sphere of a given cofactor, i. e. all atoms whose charge changes upon ET (in both heme and histidines) as these atoms constitute the inner sphere in the calculation of λ_{inner}. The final total reorganization free energy λ is then the sum of our classical λ_{outer} and the λ_{inner} of 0.05 eV.

A nonpolarizable description of electrostatics has been found to suffice for the calculation of reaction free energies [5] which is why all dynamics and per-snapshot calculations of $\frac{\partial V_\varepsilon}{\partial \varepsilon}$ for the redox potentials were done with a nonpolarizable force field. However, when it comes to reorganization free energies, nonpolarizable force fields are known to overestimate λ, as they neglect the influence of the environment's electronic polarizability on λ. [5] This can be seen by Marcus' continuum expression for λ_{outer} [50]:

$$\lambda_{outer} = (\Delta q)^2 \left(\frac{1}{\varepsilon_{op}} - \frac{1}{\varepsilon_s}\right)\left(\frac{1}{2r_1} + \frac{1}{2r_2} - \frac{1}{R}\right) \tag{2.39}$$

This equation approximates the two redox partners exchanging a charge q as spheres of radii r_1 and r_2, respectively, at a center separation of R and immersed in a dielectric medium of static dielectric constant ε_s and optical dielectric constant ε_{op}. While the latter experimentally falls in the range of 1.5-2.5 for most media [72], it is 1 with a nonpolarizable force field. Hence, the term in the central bracket (the "Pekar factor") is too large and thus is λ_{outer}. A polarizable force field provides an approximative way of explicitly incorporating electronic polarizability, thus increasing the simulated medium's ε_{op} and hence decreasing the Pekar factor in Eq. 2.39.

A comparison of nonpolarizable and polarizable force fields for ET in a cytochrome yielded a decrease by about a third when both dynamics and per-snapshot calculations of ΔE used a polarizable force field; however, most of this difference could already be attained by only calculating ΔE for each snapshot with a polarizable force field while taking the snapshots themselves from dynamics with a nonpolarizable force field. [5] In the light of these previous results, the slight increase in accuracy by running dynamics with a polarizable force field did not seem worth the large additional computational cost this would mean for all pairs in MtrF. Hence, for the calculation of λ for heme pairs in MtrF we decided to only use a polarizable force field (based on induced atomic dipoles, see Section 2.2.1) to calculate ΔE for each snapshot while taking the snapshots from nonpolarizable force field dynamics. As the nonpolarizable dynamics used in calculating the redox potentials for each heme include a window with the heme in question fully reduced (and all others oxidized), we can in fact use these fully-reduced windows for hemes i and j as the initial and final states A and B for ET from heme i to heme j - reusing their snapshots but calculating a different quantity (ΔE for ET rather than $\frac{\partial V_\varepsilon}{\partial \varepsilon}$ for oxidation). Thus, after the calculation of redox potentials, no new dynamics are required for calculating λ.

2.3 Electronic coupling matrix elements

2.3.1 Density Functional Theory

Electronic couplings need to be computed for many snapshots, restricting the acceptable computational costs of the quantum chemical method chosen. This makes the use of Density Functional Theory (DFT) appealing as DFT is generally computationally cheaper than wave function methods based on solving the Schroedinger equation [73]. DFT is based on the fact that the ground state

electron density uniquely determines the electronic Hamiltonian and thus the energy and all other properties of the system, i. e. the ground state energy is a functional of the corresponding density; furthermore, for a given external potential $V(\vec{r})$ (from e. g. a specific nuclear geometry) and fixed number of electrons the correct ground state electron density is the density that minimizes this functional. These two fundamental findings are known as the Hohenberg-Kohn theorems [74]. The energy functional $E[\rho]$ was defined as:

$$E[\rho] = \int V(\vec{r})\,\rho(\vec{r})\,d\vec{r} + F[\rho] \tag{2.40}$$

where $\rho(\vec{r})$ denotes the electron density at point $\vec{r}$ and $F[\rho]$ denotes that fraction of the total energy not due to the interaction of the electron density with the external potential; finding an actual expression for this functional is nontrivial due to the difficulty of expressing the electronic kinetic energy as a functional of the density.

A practical way of putting the energy functional $E[\rho]$ to use was provided by the Kohn-Sham approach [75] which defines an electronic wave function ψ_s for a hypothetical reference system of noninteracting electrons producing the same density as the actual system:

$$\psi_s = \frac{1}{\sqrt{N!}} \det[\phi_1 \phi_2 \cdots \phi_N] \tag{2.41}$$

where the ϕ_i are the individual spin orbitals occupied by the reference system's N electrons. The kinetic energy $T_s[\rho]$ for this refence system is then given by:

$$T_s[\rho] = \sum_i^N \left\langle \psi_s \left| -\frac{1}{2}\nabla_i^2 \right| \psi_s \right\rangle \tag{2.42}$$

This allows for $F[\rho]$ in Eq. 2.40 to be expressed as the sum of $T_s[\rho]$ and an effective one-electron potential. To this end, $F[\rho]$ is split into several terms:

$$F[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.43}$$

where $J[\rho]$, the so-called Hartree term, denotes the classic Coulomb energy of a charge density ρ:

$$J[\rho] = \frac{1}{2}\int d\vec{r}\,d\vec{r}'\,\frac{\rho(\vec{r})\,\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} \tag{2.44}$$

and the so-called exchange-correlation energy $E_{xc}[\rho]$ is given by:

$$E_{xc}[\rho] = T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho] \tag{2.45}$$

where $T[\rho]$ denotes the kinetic energy of the interacting system and $V_{ee}[\rho]$

its total electronic interaction energy, including exchange and correlation energy.

The functional derivatives $\frac{\delta J[\rho]}{\delta \rho(\vec{r})} = \int d\vec{r}' \frac{\rho(\vec{r}')}{|\vec{r}-\vec{r}'|}$ and $\frac{\delta E_{xc}[\rho]}{\delta \rho(\vec{r})} =: V_{xc}(\vec{r})$ can now be combined with the external potential $V(\vec{r})$ to yield an effective one-electron potential $V_s(\vec{r})$:

$$V_s(\vec{r}) = V(\vec{r}) + \int d\vec{r}' \frac{\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} + V_{xc}(\vec{r}) \tag{2.46}$$

This in turn allows the definition of a noninteracting Hamiltonian $\hat{\mathsf{H}}_s$ for the reference system of electrons:

$$\hat{\mathsf{H}}_s = \sum_i^N \left(-\frac{1}{2}\nabla_i^2 \right) + \sum_i^N V_s(\vec{r}_i) \tag{2.47}$$

where $\vec{r}_i$ is the position vector of the ith electron. The orbitals constituting the reference system of noninteracting electrons are then obtained as the eigenfunctions of this Hamiltonian; they are referred to as the Kohn-Sham orbitals:

$$\left(-\frac{1}{2}\nabla_i^2 + V_s(\vec{r}) \right) \psi_i = \varepsilon_i \psi_i \tag{2.48}$$

with ε_i as the ith orbital energy. As for the Hartree-Fock method, the orbitals have to be determined self-consistently. The density of this reference system and hence of the real system (since their densities were defined to be identical) is then obtained as

$$\rho(\vec{r}) = \sum_i^N \sum_\sigma |\psi_i(\vec{r},\sigma)|^2 \tag{2.49}$$

where σ denotes electron spin (α or β). Up to here, the theory is exact; however, in order to actually use it, an approximation for the exchange-correlation functional $E_{xc}[\rho]$ needs to be made. The conceptually simplest approach is to assume this functional to be an integral over a function ε_{xc} that depends on the local density [75], i. e.:

$$E_{xc}^{LDA}[\rho] = \int \rho(\vec{r}) \varepsilon_{xc}(\rho) d\vec{r} \tag{2.50}$$

Such functionals are consequently referred to as Local Density Approximation (LDA) functionals. In contrast, another class of functionals defines an ε_{xc} depending on both the local density $\rho(\vec{r})$ as well as its gradient $\nabla\rho(\vec{r})$; these are known as Generalized Gradient Approximation (GGA) functionals [76]:

$$E_{xc}^{GGA}[\rho] = \int \rho(\vec{r}) f(\rho, \nabla\rho)\, d\vec{r} \tag{2.51}$$

The GGA functionals PBE [76] and BLYP [77] have been employed in this work.

Finally, a third popular class of functionals are so-called hybrid functionals that try to improve on the exchange energy calculation in DFT by calculating the Hartree-Fock exchange energy and mixing it in (i. e., calculate the total exchange energy as a fraction from the exchange functional and another fraction from Hartree-Fock). A popular functional of this kind that is also employed in this work is B3LYP. [78, 79] Thus incorporating a fraction of "exact exchange" is computationally expensive but it does have the advantage of mitigating the self-interaction error in DFT whereby even a single electron interacts with its own charge density, thereby experiencing artificial delocalization.

As for methods based on the Schroedinger equation, it is useful to expand the Kohn-Sham orbitals in a basis set. There are two main approaches: Atom-centered basis sets that try to describe the electronic structure around atoms by a relatively small number of well-chosen localized functions; and plane wave basis sets [80] that use a vast number of plane waves of the form $f_G(\vec{r}) = \frac{1}{\Omega} \cdot e^{i\vec{G}\vec{r}}$ where $\vec{G}$ is the wave vector and Ω a cell volume; this approach assumes a periodic system with certain cell dimensions, and for cluster calculations (i. e. nonperiodic systems) the system needs to be decoupled from its periodic images (i. e., the interaction energy needs to be subtracted). In this work, plane wave calculations under cluster conditions were used. Unlike atom-centered basis functions, they describe all regions of space equally well and are hence advantageous in describing the region of space between donor and acceptor, which is crucial for the coupling matrix element calculations.

While the procedure described so far treats the system in the gas phase, it is also possible to treat the environment on a classical level, i. e. surround the quantum system with classical point charges and thus take polarization through the environment into account. In the case of MtrF, this allows us to calculate the electron densities of individual cofactors under the influence of the protein and solvent environment. This approach is known as QM/MM [81] where the "QM" stands for the quantum-mechanically treated subsystem and "MM" for the environment treated with Molecular Mechanics.

2.3.2 Constrained Density Functional Theory

The DFT approach described in the previous subsection provides an efficient means of obtaining the adiabatic ground state energy and density for a given system, but does in itself not yet yield the diabatic states ψ_a and ψ_b describing a donor/acceptor system before and after ET, which are needed to obtain the electronic coupling matrix element $H_{AB} = \langle \psi_a | \hat{H} | \psi_b \rangle$. However, the stan-

dard DFT approach can be modified so as to yield diabatic states. An approach that has been found useful for ET applications is Constrained Density Functional Theory (CDFT) where a constraint is introduced into the energy functional Eq. 2.40 to enforce a certain charge difference between donor and acceptor [52, 82]. This is achieved by adding a Lagrange multiplier V to the energy functional $E[\rho]$ to yield a new functional $W[\rho, V]$:

$$W[\rho, V] = E[\rho] + V_c[\rho, V] \tag{2.52}$$

where

$$V_c[\rho] = V\left(\int w(\vec{r})\,\rho(\vec{r})\,d\vec{r} - N_c\right) \tag{2.53}$$

Here, V_c is the constraint potential added, N_c is the desired difference in number of electrons between donor and acceptor ($+1$ or -1 in initial and final state, respectively) and $w(\vec{r})$ is a weight function defining the regions of space belonging to donor and acceptor (bearing opposite sign for these two regions). The functional $W[\rho, V]$ is then minimized to yield the lowest-energy electron density ρ that fulfils the density constraint set by the desired charge difference N_c; this density then corresponds to the initial or final diabatic state, respectively.

By enforcing localization of an excess charge onto one cofactor by means of a constraint, this procedure also mitigates the issue of the self-interaction error in DFT, i. e. the spurious interaction of an electron with its own charge density, that in the case of donor-acceptor pairs favors spurious delocalization/leakage of the excess charge from one cofactor to the other. With the excess charge constraint by the additional potential V_c, spurious delocalization is largely contained to within a cofactor.

To calculate the electronic coupling matrix element $H_{ab} = \langle \psi_a | \hat{H} | \psi_b \rangle$ with CDFT, $\hat{H}$ is approximated by the state-dependent Kohn-Sham Hamiltonian $\hat{H}^{KS}$ and the two diabatic electronic wavefunctions $\psi_{A,B}$ are approximated by the two diabatic Kohn-Sham wavefunctions that minimize $W[\rho, V]$. These are initially not yet orthogonal and thus labelled $\psi_{A,B}$ to distinguish them from the orthogonal wavefunctions $\psi_{a,b}$. In this nonorthogonal diabatic basis $\psi_{A,B}$, the electronic Hamiltonian is given as [83, 82]:

$$H' = \frac{1}{1 - S_{AB}^2} \begin{pmatrix} H_{AA} - S_{AB}H_{BA} & H_{AB} - S_{AB}H_{BB} \\ H_{BA} - S_{BA}H_{AA} & H_{BB} - S_{BA}H_{AB} \end{pmatrix} \tag{2.54}$$

Here, $H_{AA} = \langle \psi_A | \hat{H}_B^{KS} | \psi_A \rangle = E_B$ and $H_{BB} = \langle \psi_B | \hat{H}_B^{KS} | \psi_B \rangle = E_B$, i. e. these are the energies of the two diabatic states evaluated with the constraint-free Kohn-Sham Hamiltonian. H_{AB} and H_{BA} are the corresponding cross terms $\langle \psi_A | \hat{H}_B^{KS} | \psi_B \rangle$ and $\langle \psi_A | \hat{H}_A^{KS} | \psi_B \rangle$, respectively, and can be obtained from a

few other quantities resulting from a CDFT calculation [82]:

$$H_{AB} = F_B S_{AB} - V_B W_{AB} \tag{2.55}$$

$$H_{BA} = F_A S_{BA} - V_A W_{BA} \tag{2.56}$$

Here, $F_{A,B}$ are the energies of state A and B including the constraint energy; $S_{AB} = S_{BA}^*$ is the overlap between the two states; $V_{A,B}$ are the Lagrange multipliers and hence constraint energies in both states; and $W_{AB} = W_{BA}^*$ are the off-diagonal elements of the weight function matrix W whose elements are defined as $W_{AB} = \left\langle \psi_A \middle| \sum_{i=1}^{N} w(\vec{r}_i) \middle| \psi_B \right\rangle$ where $w(\vec{r})$ is the weight function used to construct the charge-constrained states. In order to obtain the desired coupling matrix elements, the Hamiltonian has to be transformed to an orthogonal basis. (For otherwise it would contain the overlap of the two diabatic states as an artifact.) This can be achieved by transforming ψ_A and ψ_B to the generalized eigenstates of the weight function matrix W; [82, 84] these eigenstates should be the states to most closely resemble the original constrained states [84] and can therefore be used as the sought-after orthonormal diabatic basis $\psi_{a,b}$. With C as the matrix of generalized eigenstates of W, the final Hamiltonian H in the proper basis is then obtained as:

$$H = C^{-1} H' C \tag{2.57}$$

and the off-diagonal elements of H constitute the final electronic coupling.

2.3.3 Fragment Orbital Density Functional Theory

The CDFT procedure outlined in the previous section is still computationally more expensive than a standard DFT calculation due to the need to converge the Lagrange multiplier as well. Therefore, another approach to obtain electronic coupling matrix elements that has been used in this work (in fact as the main method for this purpose) is Fragment Orbital Density Functional Theory (FODFT). [82] Whilst the diabatic states in CDFT are obtained by optimizing the electron density for the donor-acceptor dimer subject to a charge constraint, they are in FODFT constructed from the Kohn-Sham orbitals of the isolated fragments. I. e., an unconstrained DFT calculation is performed on the isolated donor and the isolated acceptor and a Slater determinant is formed by combining the resulting orbitals from both of these fragments (and bi-orthogonalizing them, as they are initially only orthonormal within a fragment). Hence, unlike in CDFT, polarization between the two fragments is not taken into account in FODFT; however, on top of the lesser computational

cost, this also means that any spurious delocalization between the two fragments (due to the self-interaction error) is completely avoided. Constructing this fragment-based Slater determinant in two different charge states (with the excess charge being either on the donor or acceptor) would allow for the explicit construction of the initial and final diabatic state. However, part of the FODFT approach is the assumption that the initial and final determinants only differ in their respective highest occupied molecular orbital (HOMO):

$$\psi_A = |\phi_{D,1}(\vec{r}_1)\phi_{D,2}(\vec{r}_2)\cdots\phi_{D,N+1}(\vec{r}_{N+1})\phi_{A,1}(\vec{r}_{N+2})\cdots\phi_{A,N}(\vec{r}_{2N+1})\rangle \quad (2.58)$$

$$\psi_B = |\phi_{D,1}(\vec{r}_1)\phi_{D,2}(\vec{r}_2)\cdots\phi_{D,N}(\vec{r}_N)\phi_{A,1}(\vec{r}_{N+1})\cdots\phi_{A,N+1}(\vec{r}_{2N+1})\rangle \quad (2.59)$$

with $\phi_{D/A,i}$ as the ith molecular orbital of donor or acceptor, respectively, for a system of $2N+1$ electrons. I. e., it is assumed that no significant orbital relaxation takes place upon ET. Under this assumption, the matrix element H_{AB} between the two diabatic states ψ_A, ψ_B reduces to the Kohn-Sham Hamiltonian matrix element between the two HOMOs $\phi_{D,N+1}$ and $\phi_{A,N+1}$ only:

$$H_{AB} = \left\langle \psi_A \left| \hat{H}^{KS}_{FO} \right| \psi_B \right\rangle = \left\langle \phi_D^{HOMO} \left| \hat{H}^{KS}_{FO} \right| \phi_A^{HOMO} \right\rangle \quad (2.60)$$

A third approximation is made in the construction of $\hat{H}^{KS}_{FO}$: Both donor and acceptor are calculated with $N+1$ electrons each to obtain the respective HOMO in the reduced state, $\phi_{D/A,N+1}$; when combining these two orbital sets to construct the Hamiltonian for the dimer system, the resulting Slater determinant features *both* monomers with $N+1$ electrons so that $\hat{H}^{KS}_{FO}$ is constructed for a $2N+2$ electron system.

All these assumptions render FODFT a more approximate method than CDFT; however, for the work presented in this thesis, several hundred electronic coupling matrix elements needed to be calculated. Apart from the intrinsically higher computational cost of CDFT vs. FODFT, the former also requires more manual work, e. g. when convergence difficulties for the Lagrange multiplier occur. FODFT on the other hand only requires standard ground-state DFT calculations which can hence be automatized much more easily. Thus, FODFT was used in this work to calculate coupling matrix elements for heme pairs in MtrF. In fact, when we tried to calculate CDFT couplings for a sample geometry from MtrF, we encountered severe technical issues (namely, the calculation managed to constrain the charge difference, but not the spin density, with the latter spreading to the other monomer).

However, comparative coupling calculations with CDFT and FODFT for a number of test heme dimers yielded a good agreement between the two methods (justifying the assumption made in the FODFT approach for our

systems of heme dimers) and furthermore enabled to derive a correction factor to scale FODFT-based couplings to CDFT. (See Section 3.2.2.) Hence, we could use FODFT to efficiently calculate vast numbers of individual coupling values in MtrF while accounting for the difference to the more accurate CDFT method by means of a correction scheme.

2.4 Empirical ligand docking

In the last part of this work, we were interested in studying the interaction of an OM cytochrome (specifically, MtrC) with small soluble ligands, specifically flavins. In particular, we wanted to investigate where on the cytochrome surface flavins can bind, and how strongly so. For this, we used empirical ligand docking [85]. Empirical docking generally involves studying the binding of a small (potential) ligand to a target protein by assigning a certain scoring function to a given ligand and binding pose (which may or may not correspond to an actual physical binding free energy). This is of considerable importance in drug research when new potential inhibitors to proteins of therapeutic interest are searched for.

In this work, we deployed the program Autodock [86, 87]. To determine the best binding site and pose of a given ligand on a protein, Autodock performs a global optimization of the ligand's conformation as well as position and orientation within a specified search volume containing the protein region of interest. Apart from the rigid-body degrees of freedom determining position and orientation, the ligand has internal torsional degrees of freedom that determine its conformation. I. e., torsions of aliphatic heavy-atom bonds can be chosen as torsional degrees of freedom while all bonds lengths and angles sizes are kept fixed.

The function that is optimized is a free energy function of the form [86]:

$$\begin{aligned}\Delta G = \Delta G_{vdW} \cdot \sum_{i,j}^{N} \left(\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^{6} \right) \\ + \Delta G_{hbond} \cdot \sum_{i,j}^{N} E(t) \left(\left(\frac{C_{ij}}{r_{ij}} \right)^{12} - \left(\frac{D_{ij}}{r_{ij}} \right)^{10} + E_{hbond} \right) \\ + \Delta G_{elec} \cdot \sum_{i,j}^{N} \frac{q_i q_j}{\varepsilon\left(r_{ij}\right) r_{ij}} + \Delta G_{tor} \cdot N_{tor} + \Delta G_{sol} \cdot \sum_{i_C,j}^{N} S_i V_i \cdot \exp\left(\frac{-r_{ij}^2}{2\sigma^2} \right)\end{aligned} \quad (2.61)$$

As hinted at by the subscripts, these five terms represent five separate free energy contributions: van der Waals interactions, hydrogen bonds, electrostatic interactions, torsional entropy and desolvation. The ΔG prefactors were

fitted to experimental binding affinities; thus, this function really describes a free energy, rather than a potential energy function as it is used in molecular mechanics. The indices i and j run over all pairs of ligand and protein atoms as well as all pairs of ligand atoms separated by three bonds or more. The first term (van der Waals interaction) contains a Lennard Jones 12-6 potential. The second term (hydrogen bonding) contains a directional potential including an angle-dependent weight term $E(t)$, minus the average hydrogen bonding energy E_{hbond} of a polar atom with water (i. e., the hydrogen bonding energy that is *lost* during ligand binding). The third term features a screened Coulomb potential. The fourth term (torsional entropy) is proportional to the number of torsional degrees of freedom in the ligand: It is assumed that these degrees of freedom are lost upon binding and that the amount of entropy thus lost is equal for each torsion. The last term describing desolvation of nonpolar atoms only runs over carbon atoms i_C in the ligand (and all protein atoms) and yields, for each ligand carbon i_C, a sum of fractional volumina V_j for surrounding protein atoms (the distance taken into consideration via the exponential weight factor); this is multiplied by a desolvation constant S_i for each ligand carbon.

The global optimization of ligand conformations and positions is then carried out using a Lamarckian Genetic Algorithm, i. e. a modified genetic algorithm. A genetic algorithm [88] is based on procreating a population of individuals (specified by abstract "genes" - in our case, sets of ligand internal and external coordinates) under evolutionary pressure. Thus, some fitness criterion (like the above free energy function) is used to let those individuals with higher fitness (lower binding free energy) create more offspring than those with lower fitness. In the implementation in Autodock, an initial population is created consisting of individuals with randomly determined genes (ligand coordinates). Eq. 2.61 is then used to calculate the binding free energy of each individual and hence its fitness. This serves to determine how much offspring a given individual creates; with a given frequency, crossover events exchange some coordinates between the offspring individuals and mutations unilaterally change some coordinates. The "Lamarckian" modification adds local optimization of individuals with a certain frequency, thus occasionally allowing individuals' genes to acquire adaptations to their "environment". After thus obtaining a new generation, the cycle can begin again with evaluating the fitness of this new generation of individuals. This can be continued for as many generations as seems necessary to obtain a converged binding free energy.

Typically, many (several 100) dockings are carried out for a given protein structure and ligand and the obtained optimal ligand poses from each individual run are then clustered based on RMSD agreement.

Chapter 3

Heme-to-heme electron transfer in MtrF

3.1 Free energies

3.1.1 Procedure

Setup

The crystal structure coordinates for MtrF were taken from the protein database (protein id 3PMQ). The nonpolarizable AMBER03 force field [89, 90] was chosen for the classical Molecular Dynamics, together with the TIP3P water model [91] and monovalent ion parameters from [92]. Force field parameters for the heme cofactors were taken from earlier work [71] and adapted to the cysteine linkages not present before. As mentioned in section 2.2.2, it was decided to model the electron transport as transport of negative charges rather than hole transport. Hence, nine cofactors were in the oxidized state at any time whilst the tenth one was subject to the charge morphing procedure. As the charge state with all ten hemes being oxidized was thus the common end point of each heme's Thermodynamic Integration, it was chosen as the initial charge state of the protein when the simulation was set up, serving as a starting point for the ten independent Thermodynamics Integrations. Protonation states for acidic and basic residues, including histidines not bound to a heme cofactor, were decided upon by inspection of the crystal structure as well as application of the empirical pKa estimation tool propKa 3.1 [93]. All glutamates and aspartates were taken to be deprotonated, all lysines and arginines to be protonated. From five non-heme bound histidines four were taken to be neutral (with the proton at the δ-nitrogen) and one was assumed to be protonated to the positively charged imidazolium form. One histidine's pKa value was estimated by propKa to be 6.27 with no acidic residue in its neighborhood; it was thus assumed to be neutral. Another histidine's pKa

value was estimated as 6.28; even though it it was adjacent to an aspartate, the geometry did not seem suitable for the formation of a hydrogen bond and the neutral state was also assumed here. The other two δ-protonated histidines were assigned pKa values below 5.0 by propKa, and visual inspection of the structure did not provide any reason to deviate from this assessment. The situation was different for the fifth free histidine which was assigned a pKa value of 7.74 (thus not too clearly protonated) but showed a clear hydrogen bond orientation towards a close-by aspartate. It seemed to be unambiguously protonated in the crystal structure and was thus taken to be protonated to the imidazolium form. The system was prepared using LeaP from the program suite AmberTools 1.4 [94]; after hydrogen atoms were added to the crystal structure, 35023 water molecules were added to the box in order to cover the protein in every direction with at least 15 Å of water. 85 Na and 63 Cl - ions were added to neutralize the system and reach a salt concentration of about 0.1 $mol \cdot l^{-1}$. A pH value of 7.0 was assumed. For the dynamics, periodic boundary conditions were applied and a target temperature of 300 K and a target pressure of 1.01325 bar (1 atm) were set whenever thermostats and barostats were used; all bonds including hydrogen were constrained using the SHAKE algorithm. The simulations were carried out using the molecular dynamics package NAMD 2.7 [95] on the supercomputers HECToR (based at the University of Edinburgh, UK) and Chinook (based at the Pacific Northwest National Laboratory, U.S.).

Molecular Dynamics protocol

The system was initially energy-minimized from the crystal structure for 5000 steps; subsequently the solvent was equilibrated for 500 ps (1 fs time step in the first 50 ps, 2 fs time step afterwards) with the protein kept frozen, rescaling of temperature every 5000 steps and a Langevin barostat. The protein was then slowly released, applying harmonic constraints to its atoms with decreasing force constants and still rescaling the temperature every 5000 steps while fixing the volume. This was done for force constants of 99 $kcal \cdot mol^{-1} \cdot \text{Å}^{-2}$ (500 ps, 1 fs time step; same parameters for the next three steps), 75, 50, 25, 10 (500 ps, 2 fs time step), 5 (1.0 ns, 2 fs time step from now on), 1, 0.5, 0.1, 0.05 and 0.01 $kcal \cdot mol^{-1} \cdot \text{Å}^{-2}$. After this, all restraints were dropped and the system was equilibrated with the protein fully released for 5.0 ns. From this released equilibration on, the time step was kept at 2 fs and a Langevin thermostat and barostat were used. It was then decided to have the system neutral when one heme was reduced rather than with all hemes oxidized, and a further Na^+ ion was added to the last snapshot of the released equilibration, its velocity set to zero and the system equilibrated for further 1.0 ns. The pro-

tein having one excess electron (i. e. one heme being reduced and all others oxidized) is the physical state of interest (as parameters are being calculated relating to the transfer of said electron) while the fully oxidized protein constitutes a state that does not actually occur during ET to begin with. If there has to be a net charge in some state (as the cofactor charge morphing is not compensated by introduction of another ion) it thus makes sense to not have it in the state of interest, but in the fully oxidized state.

This equilibrated system with all hemes oxidized was then taken as the starting point for the Thermodynamic Integrations. It was decided to carry out two Thermodynamic Integrations per heme, one starting from the oxidized state and morphing charges to the fully reduced state and the other starting from that reduced state and oxidizing the cofactor again. Charges were morphed in steps of 0.25 elementary charges, yielding five windows per integration with ε = 0.0, 0.25, 0.5, 0.75 and 1.0 (corresponding to the fully oxidized state). Starting from the equilibrated system, an all-oxidized window was run (to be used as the ε = 1.0 window for all ten hemes) and, in parallel, the ε = 0.75 windows for the ten hemes were started (to be then followed by the 0.5, 0.25 and 0.0 windows, the latter featuring a fully-reduced heme each). In each window, the system was equilibrated in the respective state for 2.75 ns and then a 2.75 ns production was carried out with snapshots taken in 10 ps intervals to yield 275 snapshots in total. For each subsequent window's equilibration the equilibration of the previous window provided the restart snapshot. For the integration in opposite direction, the respective 0.0 window was reused and the last snapshot of its equilibration trajectory used as the restart snapshot for the ε = 0.25 window (to be followed by an 0.5, 0.75 and 1.0 window for each heme).

Calculation of redox potentials and reorganization free energies

Eq. 2.35 was used to calculate $\frac{\partial V_{\varepsilon,elec}(\vec{R})}{\partial \varepsilon}$ for each snapshot and thus the ensemble average $\left\langle \frac{\partial V_{\varepsilon,elec}(\vec{R})}{\partial \varepsilon} \right\rangle_{\varepsilon}$ for each window. $V_{1,elec}\left(\vec{R}\right)$ and $V_{0,elec}\left(\vec{R}\right)$ were calculated using the topology files (i.e. the files containing the force field parameters) from the final and initial state, respectively, whereas $V_{\Delta,elec}\left(\vec{R}\right)$ was obtained using a topology in which all atoms in the respective heme and its histidine ligands bear a charge equal to their charge difference between initial and final state (see end of section 2.2.2) and calculating the electrostatic energy of that subsystem only. All calculations were carried out on a local workstation using a script in VMD [96] written by the author, based on the VMD plugin "namdenergy" which uses NAMD to compute electrostatic energies. The individual values of $\left\langle \frac{\partial V_{\varepsilon,elec}(\vec{R})}{\partial \varepsilon} \right\rangle_{\varepsilon}$ for each value of ε were

integrated using Composite Simpson's rule [97]:

$$\Delta G = \frac{1}{12}\left\langle \frac{\partial V_{\varepsilon,elec}\left(\vec{R}\right)}{\partial \varepsilon}\bigg|_{\varepsilon=0}\right\rangle_0 + \frac{1}{3}\left\langle \frac{\partial V_{\varepsilon,elec}\left(\vec{R}\right)}{\partial \varepsilon}\bigg|_{\varepsilon=0.25}\right\rangle_{0.25} + \frac{1}{6}\left\langle \frac{\partial V_{\varepsilon,elec}\left(\vec{R}\right)}{\partial \varepsilon}\bigg|_{\varepsilon=0.5}\right\rangle_{0.5} + \frac{1}{3}\left\langle \frac{\partial V_{\varepsilon,elec}\left(\vec{R}\right)}{\partial \varepsilon}\bigg|_{\varepsilon=0.75}\right\rangle_{0.75} + \frac{1}{12}\left\langle \frac{\partial V_{\varepsilon,elec}\left(\vec{R}\right)}{\partial \varepsilon}\bigg|_{\varepsilon=1}\right\rangle_1 \tag{3.1}$$

This yielded two integrated values for ΔG for each heme, once for the forward and once for the backward integration.

The $\varepsilon = 0.0$ windows were also used to calculate vertical electron transfer energies ΔE as required for the evaluation of reorganization free energies (Eq. 2.6). For this, the polarizable force field AMBER02 [98, 99] was employed in conjunction with the POL3 polarizable water model [100]. In lieu of a charge parametrization of the heme cofactor for the polarizable force field, the nonpolarizable charges were retained for the cofactors and their polarizabilities set to zero. The simulation program sander from the AMBER package [94] was used for all energy calculations. Outer sphere energy gaps ΔE_{outer} were calculated via Eq. 2.37 in order to calculate the outer sphere reorganization free energy λ_{outer} using Eq. 2.6.

3.1.2 Results

Fig. 3.1 shows the redox potentials obtained from Thermodynamic Integration, connected with a curve to illustrate the free energy landscape of one-electron transport through MtrF [6]. The arrangement of hemes in the plot corresponds to the arrangement in the protein (see b), with hemes 10 and 5 (left and right end in a, bottom and top in b) being the end points of the octa-heme chain and hemes 2 and 7 branching off from hemes 6 and 1, respectively. As two Thermodynamic Integrations were carried out per heme, two curves are obtained (brown and cyan). The black curve contains the average values for each heme which were taken as the final redox potentials. Half of the difference between the two individual values for each heme was used as uncertainty. (The statistical errors in the individual values were always below 10 meV and are thus not shown.) For comparison, a set of ten marks along the vertical axis describes a set of ten tentative redox potentials obtained experimentally from protein film voltammetry [2]. The redox potentials shown are relative to the standard hydrogen electrode; while the Thermodynamic Integration itself could only yield numbers containing an unknown offset (which

is equal for all ten hemes), the red set of measured values could be used as a reference towards which to uniformly shift our set of values. Thus, Fig. 3.1 contains an empirical shift of -1.567 V; however, this is irrelevant for the driving forces which are equal to the differences in redox potentials and thus are insensitive to such a uniform shift. Fig. 3.1a shows that the redox potentials are roughly symmetric with respect to the center of the protein (between hemes 1 and 6); the only major exception is the asymmetry between heme 2 and heme 7 which constitutes an outlier in the free energy landscape. However, it is suspected that the protonation state of histidine 451 (labelled "Hip451" in Fig. 3.1b) might be related to the unusually high redox potential of heme 7. This histidine was modelled as protonated due to its apparent hydrogen bond with an aspartate in the crystal structure and estimated pKa value of 7.74 (see Section 3.1.1); however, a pKa value of 7.74 would still correspond to an ensemble fraction of 15 % unprotonated histidine 451 at pH 7.0. Our method only allowed us to treat this histidine as fully protonated or fully deprotonated, so the amount of positive charge around heme 7 might in fact have been overestimated in our simulations.

The additive nature of the force field used makes it possible to split the energy gap ΔE for a given snapshot into contributions from individual residues. Averaging this contribution for a given residue over the entire window then yields its contribution to $\langle \Delta E \rangle_\varepsilon$, and integrating over all windows gives the contribution of a given residue to the redox potential in question. (Calculating ΔG in this fashion of course neglects the correction term $\left\langle V_{\Delta,elec} \right\rangle_\varepsilon$. However, this term turned out to be nearly constant between windows; hence, due to the $(2\varepsilon - 1)$ prefactor in Eq. 2.35, the correction term nearly cancels out when integrating from $\varepsilon = 0$ to 1. Correspondingly, leaving out the correction term from the integration and instead only integrating $\langle \Delta E \rangle_\varepsilon$ never changed the redox potentials by more than 1 meV.) A program to analyze energy gaps and redox potentials in this manner had been written by Dr. Jochen Blumberger and was used in this work to decompose the redox potentials obtained for the hemes in MtrF (see also Discussion in Section 3.1.3).

This residue decomposition could be used to obtain an idea of what the impact of partial deprotonation (in a time-average/ensemble sense) of histidine 451 might be: Simply scaling its contribution to the redox potential of every heme (as obtained from this decomposition) by a factor of 0.85 (i. e. considering the 15 % deprotonated histidine to not contribute) would decrease the redox potentials of heme 6, 7 and 8 while not affecting the other hemes, decreasing the overall potential range from currently 0.35 V to ca. 0.29 V. The experimentally found range was about 0.26 V [2].

The numerical values are presented in table 3.1, taking half of the differ-

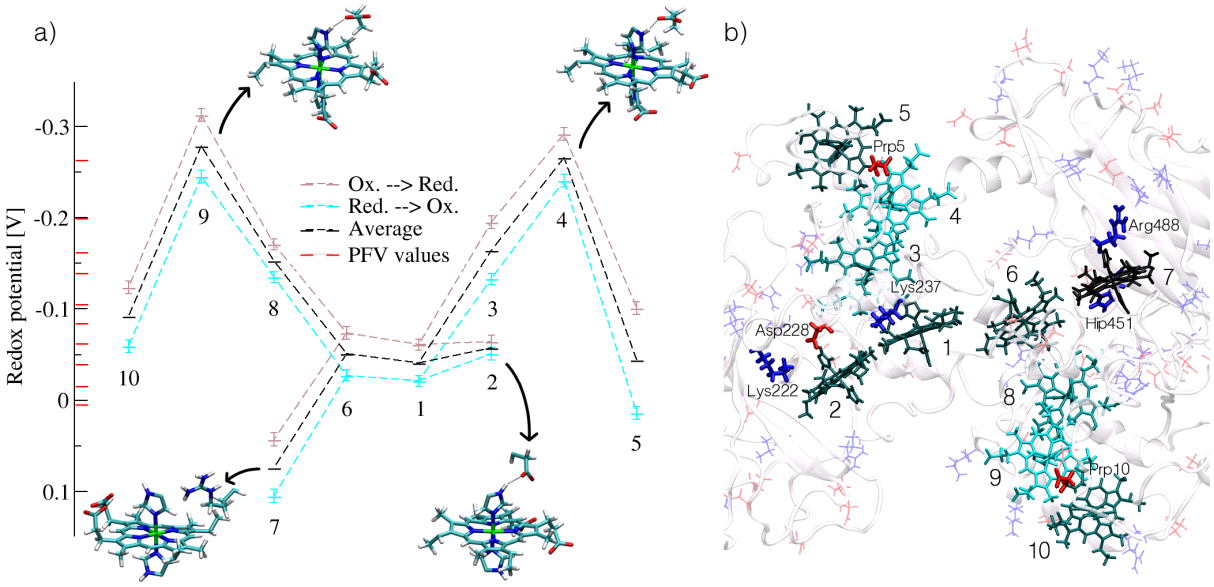

Figure 3.1: a. Redox potentials as obtained from Thermodynamic Integration (in forward and backward direction of integration, see text) for each heme. The red bars on the vertical axis denote a set of ten tentative redox potentials obtained via film voltammetry [2] which however could not be assigned to individual hemes. b. A colour map of the distribution of redox potentials in MtrF, with a lighter colour indicating a lower redox potential (and thus a higher position in the free energy landscape). Several residues of significant influence are highlighted. Reprinted with permission from [6]. Copyright 2012 American Chemical Society.

ence between the values for forward- and backward-integration as the uncertainty in the averages. In addition, table 3.2 lists driving forces for electron transfer between adjacent hemes averaged over the driving forces obtained from the brown and cyan curves in Fig. 3.1, again with half of the difference in the individual values as uncertainty. It can be seen that, due to the very similar shape of the two curves in Fig. 3.1, the uncertainty in the driving forces is generally smaller than the uncertainty in the redox potentials themselves and in most cases does not exceed ~10 % of the driving force, with the exception of pairs 1-2 and 1-6 for which the driving force almost vanishes.

As described in section 2.2.3, only the outer sphere part of the reorganization energy could be obtained here; we used an estimate of 0.05 eV for the inner sphere contribution to obtain an estimate for the total reorganization energy. The resulting numbers are shown in Table 3.3, together with the activation free energies as obtained via Eq. 2.7. Together with the driving forces, the reorganization energies then allow construction of Marcus free energy parabolas for each electron transfer step. This is shown in Fig. 3.2.

Table 3.1: Average redox potentials (vs. standard hydrogen electrode, SHE) for each heme in MtrF as obtained from Thermodynamic Integration. Half of the difference of forward- and backward-integration values was taken as uncertainty. Reprinted with permission from [6]. Copyright 2012 American Chemical Society.

Heme	Average redox potential vs. SHE with uncertainties [V]
1	-0.041 ± 0.020
2	-0.057 ± 0.007
3	-0.164 ± 0.031
4	-0.266 ± 0.026
5	-0.044 ± 0.057
6	-0.051 ± 0.024
7	0.074 ± 0.031
8	-0.153 ± 0.018
9	-0.279 ± 0.034
10	-0.092 ± 0.033

Table 3.2: Driving forces for ET between adjacent hemes in MtrF averaged over the brown and cyan curve in Fig. 3.1 with half of the respective difference as uncertainties. Reprinted with permission from [6]. Copyright 2012 American Chemical Society.

Pair (donor→acceptor)	Average driving force with uncertainty [V]
1→2	0.017 ± 0.014
1→3	0.123 ± 0.011
1→6	0.010 ± 0.004
3→4	0.103 ± 0.006
4→5	-0.223 ± 0.032
6→7	-0.125 ± 0.008
6→8	0.103 ± 0.006
8→9	0.126 ± 0.016
9→10	-0.187 ± 0.002

Table 3.3: Reorganization free energies for each heme pair in MtrF together with resulting activation free energies (as obtained from Eq. 2.7 with the driving forces ΔA from Table 3.2). Reprinted from Ref. [7].

Pair (donor→acceptor)	λ [eV]	$\Delta A^{\ddagger}$ forward [eV]	$\Delta A^{\ddagger}$ backward [eV]
1→2	1.13	0.29	0.27
1→3	0.96	0.31	0.18
1→6	0.94	0.24	0.23
3→4	0.75	0.24	0.14
4→5	0.84	0.11	0.34
6→7	1.06	0.21	0.33
6→8	0.87	0.27	0.17
8→9	0.93	0.30	0.17
9→10	0.99	0.16	0.35

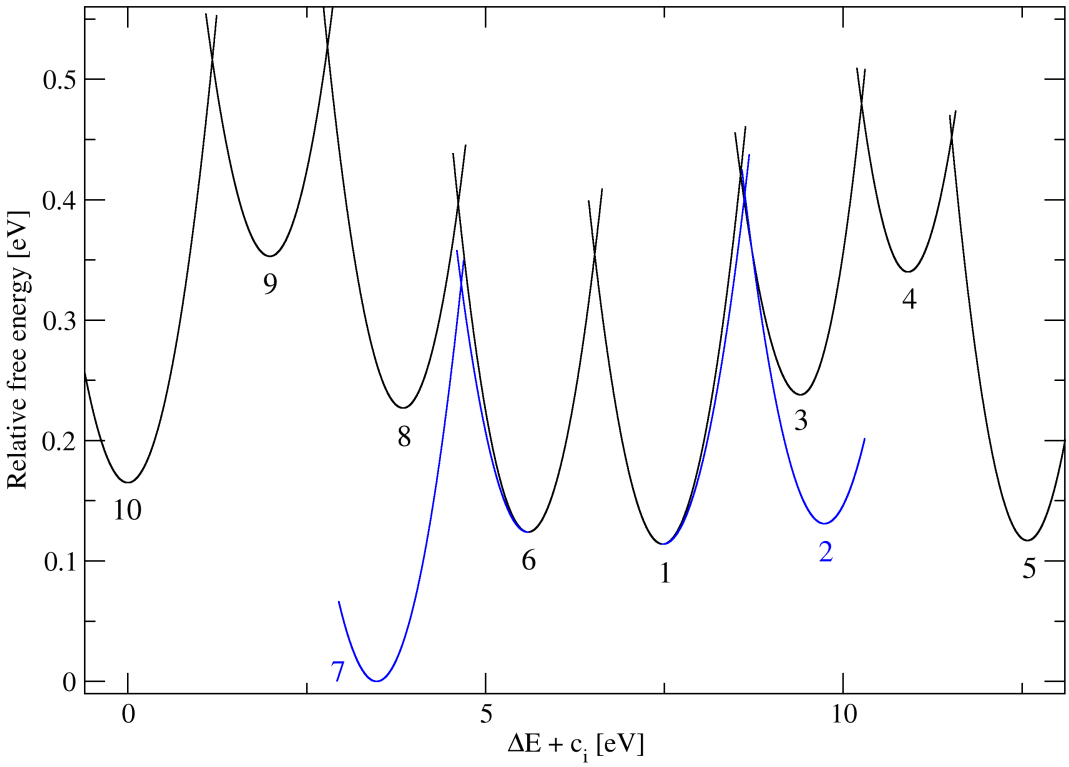

Figure 3.2: Marcus free energy parabolas constructed from the reorganization energies and driving forces determined in this work. The vertical position of the parabola minimum for each heme denotes the free energy of that heme being reduced (relative to heme 7 being reduced, which is hence at 0.0 eV) and corresponds to the heme's position in the redox potential landscape in Figure 3.1 a. From that minimum, a parabola arm rises in either direction to describe the corresponding diabatic free energy surface in ET with the adjacent heme (as in Figure 2.1). (Heme 6 and 1 each have a third parabola arm branching off towards heme 7 and 2, respectively.) The curvature of each parabola arm, as well as the horizontal spacings of parabola minima, are chosen so as to yield the reorganization energies λ and activation free energies $\Delta A^{\ddagger}$ from Table 3.3. (Hence the slightly different curvatures e. g. of the two arms branching off from heme 6 to heme 7 and 8.) ΔE on the x-axis refers to the energy gap for each pair (i. e. it is zero at the intersection of the corresponding parabola arms) and the c_i is the constant added to each pair's ΔE so as to align the parabola arms horizontally as shown. Reprinted from [7].

3.1.3 Discussion

The first interesting feature to observe about the free energy landscape is the aforementioned symmetry. This could not be expected beforehand as the cofactor arrangement is quasi-symmetric, but the distribution of charged residues (which affect the redox potentials) is not symmetric (see figure 3.1b which marks all charged residues in MtrF). Nonetheless, the two free energy hills in Fig. 3.1a have about the same height. To some extent this might be accounted for by the propionate side chains of hemes 5 and 10 that form (transient) hydrogen bonds with the histidine ligands of hemes 4 and 9 during the

dynamics. In analyzing the contributions of individual residues to the redox potentials (see Results in Section 3.1.2) it was found that these propionates are the groups that affect heme 4 and 9 the strongest. Thus, one major cause for the free energy hills at heme 4 and 9 is already present in the heme cofactor arrangement itself. Another observation in regard to the free energy profile is that the exit hemes 10, 2 and 5 feature very similar redox potentials, with the driving force from heme 10 to heme 5 amounting to ~50 meV (and even less for $10 \rightarrow 2$). This seems reasonable given that this is a mere transport process during which as little free energy as possible should be lost. The driving force for $10 \rightarrow 7$ is of course higher, but as discussed in Section 3.1.2, the high redox potential of heme 7 might be an artifact of an inaccurate treatment of the charge state of one histidine. The question remains of course why the free energy profile features uphill steps at all, rather than simply exhibiting a flat surface to maximize transport rates. A possible answer might lie in MtrF's interaction with substrates. It has been found [2] that MtrF is able to reduce the soluble electron shuttle flavin mononucleotide (FMN) whose redox potential lies at -219 mV [101]. Our data implies that only hemes 4 and 9 would be able to spontaneously reduce FMN. Hence, the free energy hills might indicate that the non-terminal hemes fulfil other functions in addition to merely moving electrons towards the terminal cofactors. This idea will be discussed further below when couplings are taken into account.

The reorganization free energies λ found fall into the range generally expected for redox proteins (0.25-1.5 eV) [55]. λ for ET between cofactors in a protein to a large extent depends on their solvent exposure [55]. The values obtained for MtrF (0.75-1.13 eV) are consistent with expectation as λ values below 0.9 eV are usually expected for solvent-shielded cofactors, with higher values typical for one solvent-exposed cofactor [55]. In MtrF, most hemes are partially solvent exposed. The not particularly large range of 0.4 eV for λ in MtrF can be rationalized with the solvent exposure being comparatively similar for most hemes (with only hemes 3 and 8 being a bit more shielded). However, no stronger correlation could be found between λ and heme solvent accessibilities.

It is interesting to note that λ already mitigates the effect of some driving forces on the activation free energies, with e. g. almost the same $\Delta A^{\ddagger}$ for $1 \rightarrow 2$ and $1 \rightarrow 3$ despite the latter featuring an uphill step in free energy. However, we still end up with activation free energies for the two energetically most unfavourable steps ($10 \rightarrow 9$ and $5 \rightarrow 4$) that are higher by ~100 meV than say for the central pair 1-6 (two basically equipotential hemes).

3.2 Electronic couplings and heme-to-heme electron transfer rates

3.2.1 Procedures

Test systems

Electronic structure of a model heme Prior to any actual coupling calculations, the electronic structure of a model heme in the gas phase (both oxidized, Fe(III), and reduced, Fe(II)) was studied with DFT, using different functionals and two different codes, the plane-wave based CPMD code [102] and NWChem [103] (used with Gaussian basis functions here). The model heme used is depicted in Fig. 3.3; it consists of the central iron ion plus a porphyrin ring with all side chains truncated and two imidazole ligands in place of the histidine ligands in the protein system. The imidazole planes in this model are parallely aligned and staggered with regard to the Fe-N bonds of the porphyrin. We optimized the geometry and calculated the orbitals using different functionals. The functionals tried were PBE (both CPMD and NWChem), BLYP (Fe(II) for CPMD, Fe(III) for NWChem) and B3LYP (both codes, but using structures optimized in NWChem). For each calculation, the nature of the highest occupied molecular orbitals and their relative energies were studied. The calculations in CPMD used a plane wave cutoff of 90 Ryd (with no subsequent single-point run with a higher cutoff, except for PBE Fe(II) - see below) and a closest atom-box distance ("vacuum gap") of 3 Å for PBE and 4 Å for BLYP and B3LYP, respectively (as the hybrid functional B3LYP required a different option to de-couple the system from its images and create cluster conditions, which in turn required a larger gap; BLYP was used to provide an initial guess for B3LYP). In NWChem the basis set 6-311++G** was used for N, C and H and Ahlrichs VTZ for Fe for all final single-point calculations. The geometry optimizations for Fe(II) used smaller basis sets for PBE (6-31G*) and B3LYP (a combination of 6-311++G** for C, N, H and 6-31G** for Fe). The Fe(III) geometry optimizations could be restarted from converged Fe(II) structures using the larger final basis set straightaway. To study the possible influence of plane wave cutoff and vacuum gap in CPMD on the electronic structure details, the orbitals for the PBE Fe(II) final geometries were recalculated once with a higher plane wave cutoff (120 Ryd instead of 90 Ryd) and once with a larger vacuum gap (4 Å instead of 3 Å).

Electronic couplings for model dimer structures After studying the electronic structure of an individual model heme in the gas phase, the CDFT and

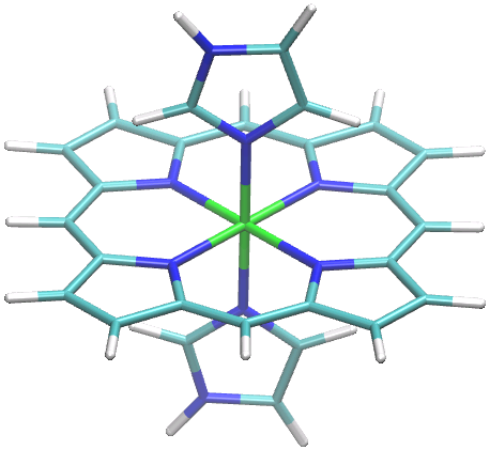

Figure 3.3: The model heme used for electronic structure investigations. Iron is shown in green, carbon in cyan, nitrogen in blue and hydrogen in white. The porphyrin ring is truncated and the histidine ligands are cut off at their imidazole ring.

FODFT method as implemented in CPMD [82] were applied to a set of model heme dimers built from the monomer structure in Fig. 4. The five model dimers (Fig. 3.4) were taken from Smith et al. [8] who also calculated electronic couplings for these dimers, thus allowing comparison of our results to theirs. In order to mimic the transition state of electron transfer, the monomer geometry used was obtained by averaging the porphyrin coordinates between BLYP-optimized Fe(II) and Fe(III) geometries and adjusting the axial Fe-N bond lengths to their average value between these geometries. The monomers were then placed at an Fe-Fe distance of 16.0 Å according to the five motifs in Fig. 4. For each dimer structure we calculated the electronic coupling once with CDFT and once with FODFT. The pseudopotentials used were Goedecker-Teter-Hutter (GTH) potentials [104] and the chosen plane wave cutoff was 130 Ryd. We tried both the PBE [76] and the BLYP functional [77]. (The PBE calculations still used the BLYP structure as we wanted to see the impact of different functionals for the same geometry.)

Sampling of coupling matrix elements

Configurational sampling via Molecular Dynamics The configurations to be used for coupling calculations should ideally be sampled from the ensemble at the intersection of the two Marcus curves in Fig. 2.1, i. e. from among the configurations with vanishing vertical energy gap ΔE. We chose to approximate this ensemble by morphing the force field charges of the two cofactors in question to the half-reduced state, i. e. the charges used for the ε = 0.5 window in the Thermodynamics Integrations, but this time on two cofactors, keeping in total one excess electron in the system. Hence, individual

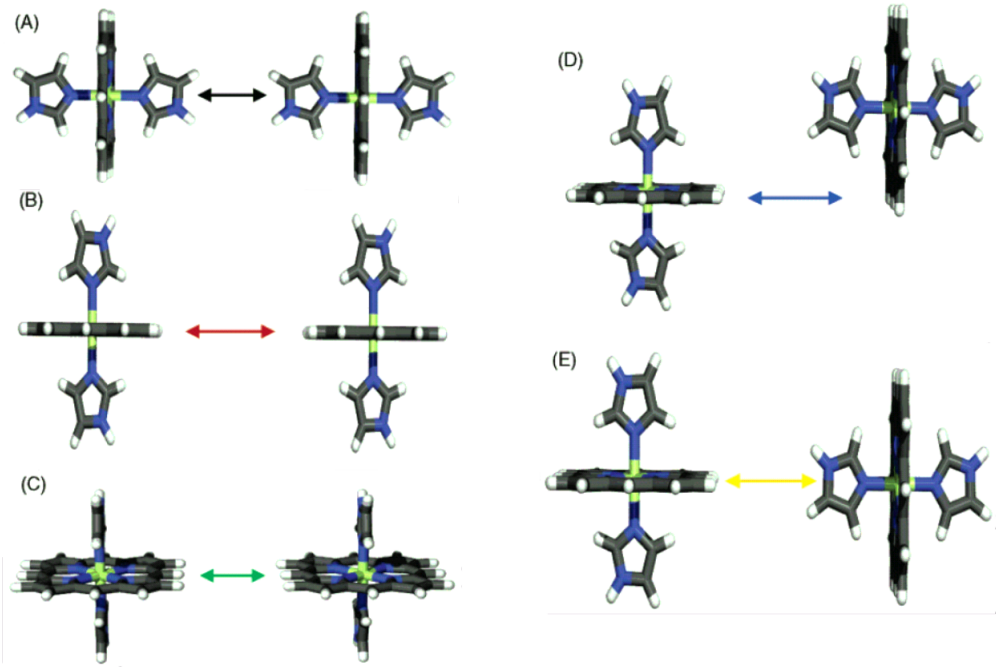

Figure 3.4: Model heme dimers of Smith et al. [8] used in this work as test systems for electronic coupling calculations (see text). Reprinted with permission from [8]. Copyright 2006 American Chemical Society.

dynamics were required for each of the nine heme pairs in MtrF. As the time scale of fluctuations in H_{AB} was not known a priori, a simulation time as long as possible was desired, subject of course to the feasibility constraint that this time had to be multiplied by nine again. A simulation time of 100 ns per pair was then chosen. The $\varepsilon = 0.5$ windows from the Thermodynamic Integrations provided restart snapshots for the half-reduced/half-reduced dynamics. In order to smooth the charge morphing, only 0.25 elementary charges were in a first step introduced into the fully oxidized partner heme and the system equilibrated for a few ns; afterwards, another 0.25 elementary charges were introduced, yielding the required charge state, and the system was equilibrated for some further ns before the production was started.

With the prospect of possibly studying time correlations of the electronic couplings, thermostat and barostat were switched off at the beginning of the production run (as no meaningful time correlations could be extracted under the influence of the stochastic Langevin thermostat). However, the energy conservation of the system had been overestimated beforehand, and the system energies increased over the course of the simulations, with a consequent rise in temperature and pressure (at most a few 10 K and few 100 bar). Upon observing this, the systems were cooled down again with a short reequilibration protocol and the dynamics were continued with thermostat and barostat switched on again. (Except for two heme pairs that had completed their 100 ns already.) No visible impact on the cofactor dynamics could be observed

for the heating trajectories.

This is not very surprising with the considerable stiffness bestowed by the c-type hemes featuring in total four bonds to the protein per cofactor (thus tightly connecting the secondary elements to which they are linked). Even though the observed increase in pressure might appear like a more significant perturbation than the temperature increase, pressure increase is in fact generally deemed a relatively mild perturbation for proteins [105] and pressures below 1 kbar are actually still considered a "normal" pressure range for proteins [106]; specifically for a c-type cytochrome (myoglobin), electron resonance studies have actually shown that distance distributions across the protein are still virtually unchanged even at 1 kbar compared to atmospheric pressure [107]. It comes therefore at no surprise that the edge-to-edge distances between the hemes were not visibly affected by the accidental perturbation and thus neither were the couplings (which are largely determined by the edge-to-edge-distance, see Section 3.2.2). Therefore, no need was seen to disregard the heated parts of the trajectories.

Couplings in the gas phase As discussed in Section 2.3.3, FODFT was chosen as the method to calculate electronic couplings for MtrF heme dimer snapshots sampled at finite temperature due to its advantages in computational cost. CDFT test calculations on some configurations for MtrF heme dimers yielded large spin leakage from the ferric to the ferrous heme, i. e. the constraint succeeded in enforcing the total charge difference but failed to constrain the spin density to the oxidized heme. However, the test calculations on the five idealized heme dimer structures (see above) showed that the FODFT couplings could be corrected with a scaling factor to reproduce the more accurate CDFT values quite well (see results in Section 3.2.3).

The FODFT protocol allows for inclusion of the protein and solvent environment via a quantum mechanics/molecular mechanics (QM/MM) approach which is available in CPMD for regular wave function optimizations [108] (which constitute the first step in an FODFT coupling calculation). In a first step though, couplings were calculated in the gas phase, i. e. only the coordinates for the two cofactors in question were used in CPMD. The functional used was PBE with a plane-wave cutoff of 130 Ryd.

25 snapshots were selected from the MD trajectory of each pair to calculate the couplings. A VMD script was written to extract the coordinates of the respective heme pair for each snapshot from the MD trajectory; in doing so saturating all carbon atoms whose bonds were cut in extracting the model system by the addition of capping-hydrogen atoms located in a distance of 1.09 Å along the severed bond. The model heme system was slightly changed compared to the test systems by including the aliphatic β-carbon of the imi-

dazole ligands as well, in order to stay consistent with previous calculations from our group [5]. The saturated extracted system was then aligned with its Fe-Fe-axis along the x-axis so as to minimize the box volume necessary to contain the system in CPMD. In order to keep the calculations as consistent as possible, the same box dimensions were used for all snapshots of a particular pair; these dimensions were determined by requiring a minimal distance of 4 Å between any box face and any atom at any time. Finding common box dimensions for all nine pairs together was deemed too inefficient as this would have required to wait for all Molecular Dynamics simulations to finish before determining the global box dimensions and being able to proceed to the DFT calculations (whereas the approach taken enabled to start with the first coupling calculations for some pairs while the Molecular Dynamics were still running for other pairs). As it was, the individual box dimensions for the nine pairs were already similar.

Couplings from QM/MM Once the 25 couplings per pair had been calculated in the gas phase, they were recalculated using orbitals optimized using QM/MM as implemented in the CPMD/Gromos interface [108]. To this end, the entire MM system was rotated and shifted for each snapshot so as to align the QM system exactly with its position in the gas phase calculation. Furthermore, as QM atoms in the CPMD implementation of QM/MM only see the central cell but not the periodic images, after rotation and shifting the MM system was rewrapped around the QM system in order to center it in the central cell, as a QM system coincidentally located on the edge of the central box would defeat the purpose of the QM/MM calculation. As the MM code Gromos only supports cubic or monoclinic cells but not the generally triclinic cell resulting from an arbitrary rotation of the MM box, the MM box dimensions had to be increased so as to describe a cuboid completely containing the rotated MM system. Thus, the MM energy in this calculation became meaningless as the periodic images were to a large part made up of vacuum. However, this was not relevant for the wave function optimization. In accordance with a previous protocol [5], the electrostatic interaction of the QM system with all MM atoms within a distance of 10 a.u. to any QM atom was calculated on a real space grid whilst MM atoms beyond that distance interacted with RESP charges fitted to the QM system. Orbital optimizations were restarted from the corresponding gas phase wave functions which worked well most of the time but not always; in problematic cases MM charges were gradually scaled up from 0.0 (gas phase) to 1.0 (full QM/MM).

For comparison, we also calculated couplings for the minimized crystal structure so as to identify the effect of thermal fluctuations. To this end, the crystal structure was energy-minimized using the same charge state for each

pair as used during the dynamics as well (i. e. the two hemes in question being half-reduced, all others oxidized). Settings were the same as for the snapshots from the dynamics (incl. the respective box dimensions).

3.2.2 Results

Electronic structure and coupling tests

Electronic structure of a single heme All functionals in both codes agreed in yielding a spin density for the oxidized cofactor (Fe(III)) shaped like an out-of-plane d-orbital perpendicular to the imidazole ligands (see Fig. 3.5); this is in agreement with the experimental evidence for a $d_{xy}^2 \left(d_{xz}, d_{yz}\right)^3$ electron configuration of low-spin ferric hemes where the out-of-plane orbitals d_{xz} and d_{yz} mix to form a set of two quasi-degenerate d_π orbitals (and d_{xy} is the in-plane d orbital.). [109] However, this unoccupied d_π orbital in the oxidized ferric heme does not correspond to what BLYP and PBE yield as HOMO for the reduced ferrous heme (Fig. 3.5): Instead, d_{xy} is given as the HOMO in both functionals with both codes, followed by d_{par} (the d_π orbital in plane with the imidazole ligands) and then d_{orth} (the d_π orbital perpendicular to the imidazole plane and corresponding to the spin density in the oxidized case, i. e. to the only singly occupied d orbital of the oxidized state). This level order was conserved, even qualitatively including the energy spacing which was roughly 120 meV or less between HOMO and HOMO-1 with near-degeneracy between HOMO-1 and HOMO-2. Increasing the box vacuum gap or plane wave cutoff for Fe(II) in CPMD with PBE (see protocol) had no significant impact.

The hybrid-functional B3LYP did yield d_{orth} as the HOMO in the reduced state, suggesting that the level reordering observed in the GGA functionals BLYP and PBE was an artifact of their more approximative nature. However, as long as this level reordering is known and persistent, it does not present an obstacle for calculating couplings via the FODFT approach: All that needs to be known is which orbital is occupied in the reduced state but not in the oxidized one, i. e. which orbital constitutes $\phi_{D,N+1}(\vec{r}_{N+1})$ in Eq. 2.58. The results presented in the next section show that couplings obtained with CDFT can be reproduced well with FODFT and a GGA functional provided that the level reordering is taken into consideration.

Electronic couplings for model dimers Table 3.4 shows the coupling matrix elements obtained for the model dimers A-E from figure 3.4. The upper half shows the couplings from CDFT (PBE) and for HOMO, HOMO-1 and HOMO-2 from FODFT with PBE, together with the ratio CDFT/FODFT cou-

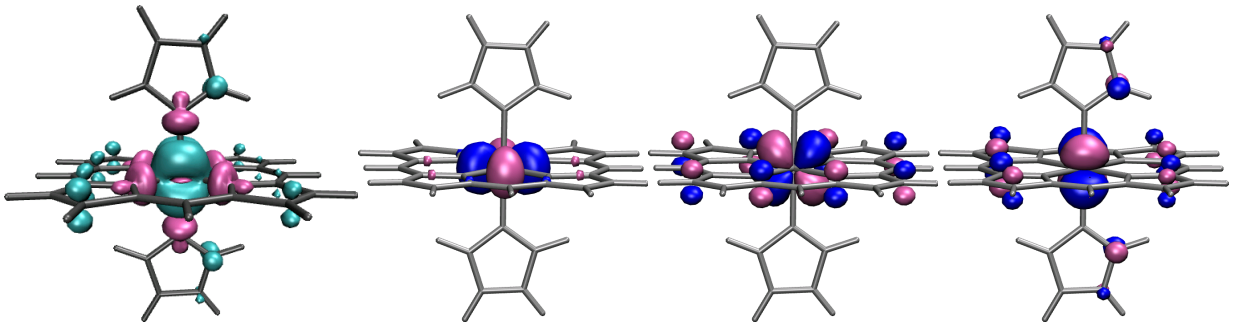

Figure 3.5: Electronic structure of a single model heme. Left: Fe(III) spin density; following: HOMO, HOMO-1 and HOMO-2 as obtained with the functionals PBE and BLYP for Fe(II).

plings (using HOMO-2 for FODFT). The lower half shows the couplings by Smith et al. [8] (scaled using their decay constants as their Fe-Fe distances were higher by 0.5 Å) and FODFT couplings with BLYP.

Table 3.4: Electronic coupling matrix elements as obtained for model dimers A-E with different methods. "Smith" refers to the values by Smith et al. [8] HOMO, HOMO-1 and HOMO-2 refer to the corresponding FODFT couplings for these orbitals. The last column shows the ratios of CDFT couplings to the FODFT (PBE) values for HOMO-2.

Dimer	CDFT (PBE) [meV]	FODFT (PBE) [meV]			CDFT/FODFT
		HOMO	HOMO-1	HOMO-2	
A	0.21	0.27	0.11	0.10	2.12
B	0.79	0.70	0.64	0.48	1.65
C	0.97	0.21	0.10	0.51	1.91
D	0.16	0.21	0.10	0.10	1.57
E	0.01	0.14	0.51	0.01	1.45
Dimer	Smith [meV]	FODFT (BLYP) [meV]			
		HOMO	HOMO-1	HOMO-2	
A	0.33	0.26	0.10	0.10	
B	1.90	0.69	0.48	0.65	
C	1.42	0.20	0.51	0.10	
D	182	0.21	0.11	0.10	
E	0.03	0.14	0.01	0.51	

Comparison of the CDFT data in Table 3.4 with the numbers by Smith et al. shows the CDFT numbers to be lower by a factor of 1.5-2 for structures A-C, which would translate to a rate difference by a factor of 2-4 (as the rate is determined by $|H_{ab}|^2$), i. e. the disagreement between the two data sets still lies within chemical accuracy. This is interesting to note as they are based on different approaches: Smith et al. used atom-centered basis sets to construct their quasi-diabatic wave functions, without any charge constraints. The relative difference is a little larger for structure E but these values (around 0.01 meV) are extremely small to begin with. The values for this structure do agree with the trend observed for structures A-C, namely that the CDFT val-

ues are consistently slightly smaller. In particular, this means that the two methods also largely agree about relative trends, with structures B and C featuring the highest couplings, structure A being an order of magnitude smaller and structure E being smaller by another order of magnitude. The only actual disagreement between the two data sets can be found for structure D; however, this structure is a clear outlier (by two orders of magnitude) even within the data set of Smith et al. Hence, our coupling value for this structure seems somewhat more plausible than the Smith value. For the other four structures, we observe the aforementioned agreement for numbers spanning two orders of magnitude.

While in CDFT the couplings are calculated between Kohn-Sham determinants, in FODFT couplings are obtained between single Kohn-Sham orbitals, specifically the HOMOs of the reduced donor and acceptor, respectively. However, as noted before, we found GGA functionals to yield a different d-orbital order from what experimental spectroscopic data implies. [109] Therefore, Table 3.4 lists the couplings for the three highest occupied molecular orbitals for PBE and BLYP, respectively. As can be seen, for PBE only the HOMO-2 couplings reproduce the trends observed with CDFT. This is consistent with the observation that, as already illustrated in Fig. 3.5, the d-orbital that, according to the spin density, is singly vacant in the oxidized heme in fact becomes the HOMO-2 in the reduced heme with PBE. Consequently, it is this orbital that constitutes $\phi_{D,N+1}(\vec{r}_{N+1})$ in Eq. 2.58, i. e. the orbital whose occupancy changes during ET, and whose one-orbital coupling matrix element can be taken as the FODFT approximation to the whole-determinant coupling.

It can be seen that the PBE-FODFT couplings are consistently smaller than their CDFT counterparts which can be rationalized by the missing polarization influence between the two dimers in FODFT. Fortunately, the ratio between CDFT and FODFT numbers falls within a comparatively narrow range of 1.5-2.2 (see last column in Table 3.4), with an average of 1.75. Hence, applying this scaling factor to all PBE-FODFT couplings could be expected to approximately correct for the missing polarization in FODFT and this approach will be taken when calculating the couplings for heme pairs in MtrF.

Lastly, a comparison of the FODFT couplings obtained with PBE and BLYP shows that these are virtually identical - except for an exchange of HOMO-1 and HOMO-2 couplings. This can be explained by the fact that BLYP simply ordered the d_π-orbitals the other way around (with d_{orth} being the HOMO-1).

Sampling of couplings

At first, all couplings for heme pairs in the protein were calculated in the gas phase, as described in Section 3.2.1. In the case of the five model dimers, the unoccupied d-orbital in the oxidized heme could be clearly identified as the HOMO-2 in the reduced heme (d_{orth}, i. e. the d_π-orbital whose lobes lie perpendicular to the plane of the two histidines), consistent with the HOMO-2 couplings showing the best agreement with the CDFT numbers. The hemes in MtrF however feature histidines at arbitrary angles with respect to each other and hence the distinction between perpendicular (d_{orth}) and parallel (d_{par}) d_π-orbitals becomes impossible. These two orbitals (which were still HOMO-1 and HOMO-2 for the geometries taken from MtrF) are quasi-degenerate and we therefore decided to follow a convention [83] for degenerate orbitals to use the RMS average of all possible couplings (i. e., between the HOMO-1 orbitals, between the HOMO-2 orbitals and the two cross-couplings):

$$|H_{\mathrm{ab}}| = c_{corr} \cdot \sqrt{\frac{1}{4} \sum_{i,j=1,2} \left\langle d^{\mathrm{A}}_{\pi,i} \middle| \hat{H}^{KS}_{FO} \middle| d^{\mathrm{D}}_{\pi,j} \right\rangle^2}, \tag{3.2}$$

where $d_{\pi,1}$ and $d_{\pi,2}$ are the two d_π orbitals on each heme (D for donor, A for acceptor) and $\hat{H}^{KS}_{FO}$ denotes the Kohn-Sham Hamiltonian constructed from the separate donor and acceptor fragments, as discussed in Section 2.3.3. c_{corr} is the aforementioned correction factor of 1.75 to account for the missing polarization from the other cofactor, as derived by comparing FODFT couplings to CDFT values (see previous section). It should be noted that the ensemble averages of the four individual couplings are similar to the ensemble average of the RMS averaged coupling.

The results for all couplings in MtrF are summarized in Table 3.5 and Fig. 3.6. Table 3.5 lists the couplings obtained for each heme pair - both the single coupling values $|H_{\mathrm{ab}}|$ from the minimized crystal structures and the ensemble RMS averages $\langle |H_{\mathrm{ab}}|^2 \rangle^{\frac{1}{2}}$ over all 25 snapshots from finite-temperature dynamics, calculated in gas phase and in the protein environment using QM/MM conditions. (For pair 1-2, the QM/MM couplings for five snapshots were ignored as the wave function optimizations with QM/MM yielded spurious electron density accumulation on a positively charged protein residue from the MM region; this meant a spurious coupling contribution, resulting in coupling values up to a factor of 30 higher than the average of the unaffected couplings.) The couplings are well converged with 25 snapshots: If we leave out every second snapshot, $\langle |H_{\mathrm{ab}}|^2 \rangle^{\frac{1}{2}}$ changes by at most 10 %. The last three columns in Table 3.5 show the edge-to-edge distance in the minimized crystal structures as well as averaged over the MD snapshots, together with their σ. (The edge-to-edge distance is calculated as the closest distance between any

Table 3.5: Coupling matrix elements $|H_{ab}|$ obtained for heme pairs in MtrF, together with edge-to-edge distances between the respective hemes. MD denotes data from thermal averaging, CS denotes data from a minimized crystal structure. Table adapted from [4].

pair $i-j$	$\lvert H_{ab}\rvert$ (CS, QM/MM) [meV]	$\left\langle \lvert H_{ab}\rvert^2 \right\rangle^{\frac{1}{2}}$ (MD, gas phase) [meV]	$\left\langle \lvert H_{ab}\rvert^2 \right\rangle^{\frac{1}{2}}$ (MD, QM/MM) [meV]	$\sigma(\lvert H_{ab}\rvert)$ (MD, QM/MM) [meV]	R (CS) [Å]	$\langle R\rangle$ (MD) [Å]	$\sigma(R)$ (MD) [Å]
1-2	0.27	0.23	0.24	0.09	6.26	6.98	0.54
1-3	0.31	0.47	0.49	0.21	6.17	5.95	0.15
1-6	0.08	0.14	0.13	0.05	6.93	6.78	0.47
3-4	1.71	2.10	2.21	0.67	4.24	4.28	0.18
4-5	2.34	3.67	3.63	1.26	3.87	3.83	0.24
6-7	0.28	0.23	0.23	0.11	6.03	6.20	0.41
6-8	0.29	0.31	0.31	0.16	5.94	5.85	0.24
8-9	2.64	2.18	2.31	1.08	4.42	4.42	0.27
9-10	4.59	4.36	4.52	1.52	3.94	3.82	0.17

two carbon atoms of the two porphyrin rings.) First considering the effect of the QM/MM environment, we see by comparing the couplings for each pair in columns 3 and 4 that the impact of the protein environment is small, with the gas phase numbers deviating less than 10 % from the corresponding QM/MM values. This suggests that the polarizing influence of the environment on the coupling orbitals is very small. Comparing the RMS averages from MD to the crystal structure couplings, the impact is a bit larger with a factor of up to ~1.6. In terms of ET rates (determined by $\langle|H_{ab}|^2\rangle$), this translates to a difference by a factor of less than 3 which, given that biological ET rates span many orders of magnitude, is still comparatively little. Fig. 3.6 shows that the individual couplings $|H_{ab}|$ are mainly determined by the edge-to-edge distance (see below for the detailed discussion); thus, the relatively small impact of thermal fluctuations on $\langle|H_{ab}|^2\rangle^{\frac{1}{2}}$ can be rationalized by comparing these edge-to-edge distances for each pair between crystal structure and dynamics (last three columns in Table 3.5): Apart from the difference of 0.7 Å for pair 1-2, the crystal structure and dynamics values are virtually identical. In addition, the fluctuations in the edge-to-edge distances are small as can be seen from the standard deviations in the last column (consistent with the significant stiffness bestowed by the covalently bound hemes); correspondingly, the fluctuations in the couplings (see $\sigma(|H_{ab}|)$) are comparatively small as well (never exceeding half of $\langle|H_{ab}|^2\rangle^{\frac{1}{2}}$).

Fig. 3.6 shows the individual (QM/MM) couplings for all snapshots vs. the respective edge-to-edge distance. The left y-axis shows the couplings on a logarithmic scale and the opposite y-axis denotes the corresponding maximal

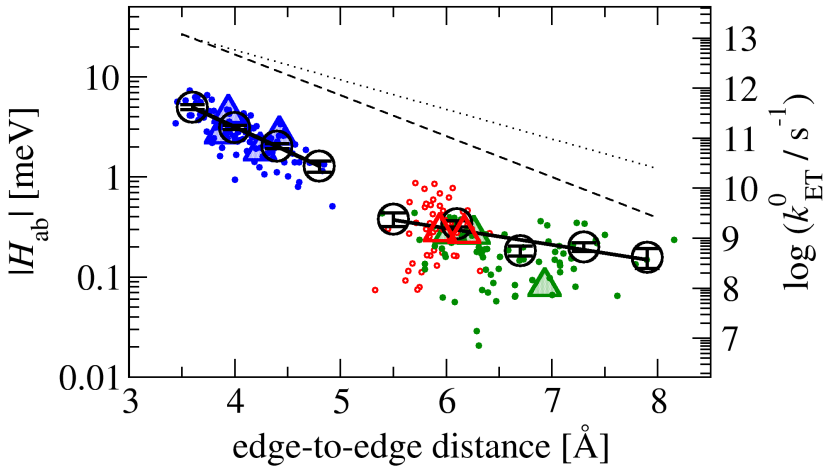

Figure 3.6: Modulus of electronic coupling matrix elements ($|H_{ab}|$) vs. heme edge-to-edge distance for all snapshots and heme pairs in MtrF (couplings incl. QM/MM). Points are colored according to the heme-heme orientation: stacked in blue, T-shaped in hollow red and coplanar in green. (See Figure 1.6 C for a depiction of the three motifs.) $|H_{ab}|$ values obtained for the crystal structure configuration are indicated by triangles (four for stacked, two hollow ones for T-shaped, three for coplanar). Root-mean-square averages of the scattered data points, $\langle|H_{ab}|^2\rangle^{\frac{1}{2}}$, were calculated for bins (denoted by black circles) of width 0.4 Å (left set) and 0.6 Å (right set) and fit to two separate exponentials, one for the stacked heme pairs and one for the T-shaped/coplanar heme pairs (solid black lines). Corresponding free energy-optimized ET-rates k^0_{ET} ($\lambda = -\Delta A$) are indicated on the axis to the right. The empirical 'Moser-Dutton ruler' (see text) is shown for the default packing density ($\rho = 0.76$, dotted lines) and for a reduced packing density accounting for through space tunneling ($\rho = 0.48$ [9], dashed lines). Figure reprinted from Ref. [4].

ET rate (assuming $\Delta G = -\lambda$). The small, coloured circles denote individual coupling points and the large, hollow circles show RMS averages for bins of 0.4 Å (left) and 0.6 Å (right) width, respectively. The colour code is based on the three different heme dimer motifs present in MtrF (see discussion in Section 1.2): Stacked pairs (3-4, 4-5, 8-9 and 9-10) are shown in blue, T-shaped pairs (1-3 and 6-8) in red and coplanar pairs (1-2, 1-6 and 6-7) in green. The nine hollow triangles (in the same colour scheme) depict couplings for the minimized crystal structures.

The stacked pairs exhibit the smallest distances (3.6-5.0 Å), followed by the T-shaped (5.2-6.5 Å) and coplanar motif (5.3-8.3 Å). The distance dependence of the resultant bin RMS averages can be best described by two exponential decay constants, one for the stacked motif, $\beta = 2.25$ Å^{-1} and

$A = 5.55$ meV ($R^2 = 0.9997$), and one for the T-shaped/coplanar motifs, $\beta = 0.8$ Å^{-1} and $A = 0.8$ meV ($R^2 = 0.85$), using the fit function $\langle |H_{ab}|^2 \rangle^{1/2}(r) = A\exp[-\beta(r-r_0)/2]$, $r_0 = 3.6$ Å (van der Waals contact). Although a fit of all nine bin averages to a single exponential with parameters $\beta = 1.65$ Å^{-1} and $A = 3.77$ meV still yields a reasonable correlation of $R^2 = 0.91$, two separate fits clearly give a better description of the data. The same fitting procedure using the Fe-Fe distance rather than edge-to-edge distance would yield $\beta = 0.6$ Å^{-1} ($R^2 = 0.79$) for the T-shaped and coplanar pairs (i. e., similar to the 0.8 Å^{-1} with ($R^2 = 0.85$) when edge-to-edge is used) and $\beta = 1.05$ Å^{-1} with $R^2 = 0.96$ for the stacked pairs. The increase in β can easily be rationalized by the fact that for the stacked pairs, the Fe-Fe distance increases about twice as fast as the edge-to-edge distance (i.e. for an edge-to-edge distance increase of 1 Å the Fe-Fe distance increases by around 2 Å) so that the distance decay is just about half as strong. The much better correlation coefficient R^2 for edge-to-edge (0.9997) however suggests that the latter is a better descriptor for the stacked pairs. This can be explained by the shape of the coupling orbitals: As can be seen for the ideal heme monomer in Fig. 3.5, the d_π-orbitals actually feature minor lobes on the porphyrin ring as well. The greater correlation of couplings with edge-to-edge distance then suggests that these minor lobes actually determine the overall coupling to a large extent. A global fit for the Fe-Fe distance would yield a decay constant of $\beta = 1.30$ Å^{-1} ($R^2 = 0.90$), slightly smaller than the 1.65 Å^{-1} from above for edge-to-edge; however, the two individual fits describe the individual bin points better as measured by average absolute differences between bin points and regression.

The decay constants obtained for the different motifs are within, or close to, the range that Smith et al. obtained when they estimated the distance dependence of the couplings for their five model dimers in the gas phase [8], namely 1.1-2.6 Å^{-1} (excluding the outlier structure D). Comparing the global fit β values to experimental estimates for tunneling through proteins, we observe that our numbers are slightly higher: 1.65 Å^{-1} vs. 1.4 Å^{-1} using the edge-to-edge metric [110] and 1.30 Å^{-1} vs. 1.1 Å^{-1} using the Fe-Fe distance [17, 14]. However, our couplings are based on through-space tunneling between closely spaced cofactors, whereas the experimental data are predominantly based on protein-mediated electron tunneling reactions over significantly longer distances. Interestingly, the free-energy optimized tunneling rates corresponding to the calculated couplings (k_{ET}^0, $\Delta A = -\lambda$) are 1-2 orders of magnitude below the empirical "Moser-Dutton ruler" (dashed and dotted black lines in Fig. 3.6) [110]. This "ruler" is given by the equation (for optimal driving forces, $\Delta A = -\lambda$) [110]:

$$log\ k_{ET}^{0} = 13 - (1.2 - 0.8\rho)(R - 3.6) \tag{3.3}$$

In this model, the protein matrix is described as a homogeneous square potential barrier to electron tunneling between two cofactors with an edge-to-edge distance R (in Å), the barrier height being determined by the effective protein packing density ρ. A general value of 0.76 has been suggested for this packing density [110] (dotted line in Fig. 3.6) while for closely spaced cofactors a reduced effective density of 0.48 has been suggested [9] (dashed line in Fig. 3.6). However, for very close distances like those in MtrF, the Moser-Dutton ruler relies on a few known rates for bacterial reaction center and photosystem proteins; comparing crystal structures shows that despite similar edge-to-edge distances, the porphyrin overlap in the special pair of the reaction center protein is much larger than in MtrF (Fig. 3.7). Hence, if the empirical ruler is fitted so as to describe rates in the photosynthetic reaction center, it can be expected to overestimate rates for heme dimers in MtrF. Hence, our findings emphasize the importance of specific local molecular structure in understanding multi-heme ET kinetics beyond empirical rules.

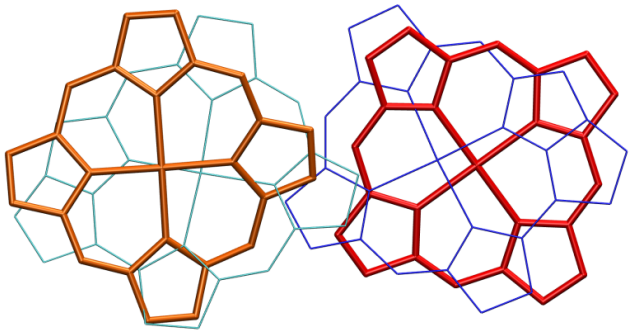

Figure 3.7: Comparison of heme stacking in MtrF and a bacterial reaction center: The stacked pair 4-5 from MtrF is shown in red/orange and the chlorophyll-special pair from the photosynthetic reaction center from *Rh. sphaeroides* (pdb code 1M3X) in blue/cyan. In spite of a similar edge-to-edge distance, the porphyrin overlap is visibly larger in the case of the special pair. Figure adapted from Ref. [4].

The ensemble-averaged couplings $\langle |H_{ab}|^2 \rangle^{\frac{1}{2}}$ (from QM/MM) can now be combined with the previously obtained driving forces ΔA and reorganization free energies λ to calculate heme-to-heme k_{ET} rates for all heme pairs in MtrF via the nonadiabatic Marcus rate equation, Eq. 2.1. Table 3.6 summarizes the three individual quantities (couplings, λ and ΔA) for each pair and lists the resulting ET rates k_{ji} and k_{ij} for ET from heme i to heme j or vice versa, respectively. Comparing the couplings (a few meV at most) to the reorganization energies λ (in the range of 1 eV), we see that $|H_{ab}| << \lambda$, i. e. ET between the hemes can be safely assumed to take place in the nonadiabatic

Table 3.6: Quantities entering the Marcus rate equation Eq. 2.1 for each heme pair in MtrF: Couplings $\langle|H_{ab}|^2\rangle^{\frac{1}{2}}$ (from QM/MM), driving forces ΔA_{ji} [6], and reorganization free energies λ [7]. The last two columns contain the resulting heme-to-heme ET rates, k_{ji} and k_{ij}. The notation ΔA_{ji} and k_{ji} refers to ET from heme i to j and k_{ij} is for ET from heme j to i. Table adapted from Ref. [4].

pair $i-j$	$\left\langle\|H_{ab}\|^2\right\rangle^{\frac{1}{2}}$ [meV]	ΔA_{ji} [eV]	λ [eV]	k_{ji} [s^{-1}]	k_{ij} [s^{-1}]
1-2	0.24	0.02	1.13	1.18×10^4	2.19×10^4
1-3	0.49	0.12	0.96	2.89×10^4	3.37×10^6
1-6	0.13	0.01	0.94	2.68×10^4	3.95×10^4
3-4	2.21	0.10	0.75	8.10×10^6	4.19×10^8
4-5	3.63	-0.22	0.84	3.12×10^9	5.81×10^5
6-7	0.23	-0.13	1.06	2.99×10^5	2.38×10^3
6-8	0.31	0.10	0.87	4.47×10^4	2.31×10^6
8-9	2.31	0.13	0.93	8.46×10^5	1.11×10^8
9-10	4.52	-0.19	0.99	6.14×10^8	4.43×10^5

regime.

The resulting rates for each pair in both directions ($i \rightarrow j$ and $j \rightarrow i$) are also compared in Fig. 3.8A. It becomes apparent that the steps that feature the greatest energetic uphill steps (namely, $10 \rightarrow 9$ and $5 \rightarrow 4$) do in fact not have the lowest rates; instead, they are still an order of magnitude higher than most of the rates along the central tetra-heme plain. Panel B provides an explanation of this: Here, the RMS couplings $\langle|H_{ab}|^2\rangle^{\frac{1}{2}}$ (motif colour code same as in Fig. 3.6) are superimposed onto the free energy landscape introduced in Section 3.1.2. As can be seen, the same steps that feature the greatest energetic uphill steps also have the highest couplings, whereas the coplanar pairs in the center (green circles) feature the smallest couplings. The impact of the difference in couplings is considerable: The order of magnitude difference between highest and lowest couplings corresponds to two orders of magnitude difference in the pre-exponential term of the Marcus rate equation (which contains $\langle|H_{ab}|^2\rangle$). Accordingly, without this correlation of couplings and driving forces (i. e., if e. g. the same nine couplings were distributed in a different manner across the nine pairs), the two energetic uphill steps $10 \rightarrow 9$ and $5 \rightarrow 4$ could become kinetic bottlenecks as well. This aspect is further explored in the next chapter. Another aspect to note in Fig. 3.8 is the (approximate) symmetry in the couplings (on top of the previously discussed free energy symmetry) and hence in the calculated ET rates, suggesting that electron conduction through MtrF is equally fast along either direction of the octa-heme chain.

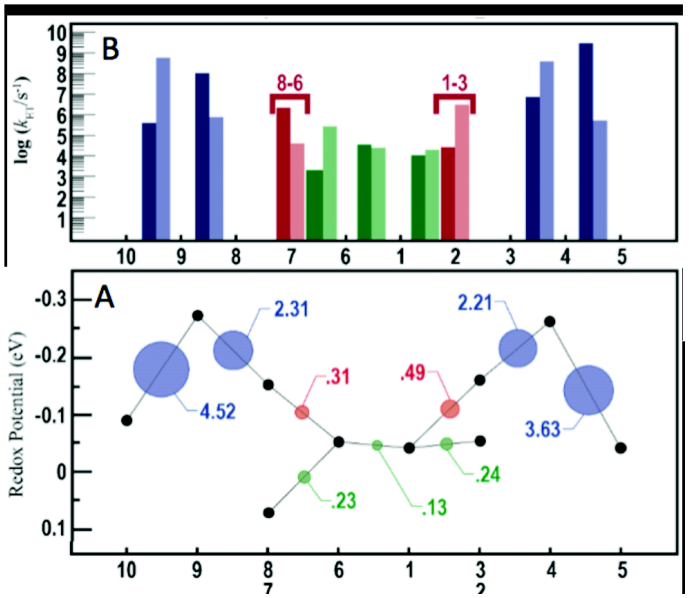

Figure 3.8: Kinetics of ET through MtrF and superposition of constituent quantities. (A) Individual ET rates k_{ET} for each pair in forward (left $\rightarrow$ right, dark bars) and backward direction (right $\rightarrow$ left, light bars) as obtained via Eq. 2.1 (see also Table 3.6). (B) The free energy landscape for ET through MtrF (see Section 3.1.2 and Figure 3.1 a) together with the RMS coupling $\langle |H_{\mathrm{ab}}|^2 \rangle^{\frac{1}{2}}$ for each pair (circles, area proportional to the coupling). The color code of the circles corresponds to the three heme pair motifs as in Fig. 3.6. Figure adapted from Ref. [4]. (Figure prepared in collaboration with Cortland Johnson.)

3.2.3 Discussion

The most significant outcome of our calculation on $\langle |H_{\mathrm{ab}}|^2 \rangle^{\frac{1}{2}}$ for each heme pair in MtrF is the aforementioned correlation with the free energy landscape as illustrated in Fig. 3.8, enabling the uphill ET rates at these heme pairs to compete with (exceed, actually) the rates among the equipotential central coplanar hemes. Hence, the low-potential hemes 4 and 9 can be incorporated without creating kinetic bottle necks. In the next chapter we study in more detail the importance of this matching between electronic couplings and driving forces for the overall transport of electrons through MtrF.

Here, we would like to raise another question in regard to the heme pair motifs occuring in MtrF: Given that stacked heme pairs exhibit the highest couplings, why does the protein feature non-stacked heme pairs at all? Three issues seem relevant here. Firstly, a higher total flux may not yield any metabolic benefit if the preceding metabolic reactions or the electron output to external substrates is rate limiting. In this case there is no evolutionary pressure to increase the flux by further optimizing the mutual orientations of all heme pairs. Secondly, the coplanar pairs may serve a distinct function in addition to electron transport along the heme network. The coplanar tetra-heme chain 2-1-6-7 exhibits a large contiguous surface area, which could possibly

form an effective multi-point contact site for solid substrates that would still be accessible if hemes 10 and 5 were already docking to substrates or partner cytochromes. Structural modeling suggests that it would be difficult to obtain the same contiguous area if hemes 1 and 6 were part of a continuous stacked octa-heme chain. Thirdly, non-stacked hemes are necessary for the formation of the staggered cross protein motif. The latter could serve as a building block for a supramolecular 2-dimensional network as previously assumed in a modeling study of a conducting bacterial pilus [111].

Further conclusions can be drawn in regard to protein design. According to our calculations, the impact of the protein electrostatic environment on the heme-heme couplings is practically negligible (difference < 10 %); that is, under through-space tunneling the couplings are solely determined by the heme-heme packing motif and interheme distances. This has the remarkable implication that as long as ET takes place via through-space tunneling without protein mediation, electronic couplings and driving forces could be modulated independently from each other (both during protein evolution and, hypothetically, in any artificial protein redesign).

In particular, this independence of the couplings from the surrounding protein suggests that the coupling results obtained for MtrF can be transferred to its homologues MtrC, OmcA and UndA that do not only share the same basic heme arrangement motif (the staggered cross) but even feature the same heme pair motifs (stacked, T-shaped and coplanar) for most of their analogous heme pairs. In addition, both MtrC and UndA feature the same proximity between propionate side chains of the outermost hemes of the octa-heme chain (hemes 5 and 10 in MtrF/C) and the histidines of the adjacent hemes (4 and 9 in MtrF/C) which we found to be a major factor in decreasing the redox potential these hemes in MtrF, respectively (see Section 3.1.3). (In UndA, heme 4 does not have a propionate closeby but some other negative residue.) OmcA still features this propionate-histidine motif for one heme. Thus, these homologues might feature the same kind of barriers in their free energy profiles as MtrF, with the overall profile and rates potentially being modulated with respect to MtrF through a different protein environment. The same correlation between driving forces and couplings that we found for MtrF might then to a certain extent apply to these other proteins as well, suggesting a design principle of wider significance.

Chapter 4

Through-protein electron transport in MtrF

In this chapter we combine the k_{ET} rate constants for heme-to-heme ET in MtrF to describe electron transport through the entire protein. For this, we consider stepwise ET along a linear chain of hemes in MtrF as well as heterogeneous ET between the first and last heme of that chain and the environment. By assuming different boundary conditions we seek to model several experimental setups. Specifically, we are interested in modelling electron flux through MtrF and the trans-membrane complex MtrCAB under physiological conditions, namely in an experimental *in vitro* setup designed to mimic *in vivo* conditions [10]; and in describing the current I through MtrF as a function of an external bias potential V in the context of a tunneling spectroscopy setup [48]. The formalism we deploy in both cases is a steady-state formalism [112, 113] that yields the electron flux J under the condition that the average population on each site (i. e., heme along the chain) is constant (as is the net flux between any two sites).

Specifically, it is assumed that the electron flux from heme i to heme j, $J_{ji} = J_{j \leftarrow i}$, can be described by a master equation

$$J_{ji} = k_{ji} P_i (1 - P_j) - k_{ij} P_j (1 - P_i) \tag{4.1}$$

with k_{ji} the ET rate constant according to Eq. 2.1 and P_i the electron population of heme i (0 for oxidized heme, 1 for reduced) [112, 113]. The terms $(1 - P_i)$ account for the fact that each heme can be occupied by only one excess electron (and thus can receive an electron in the oxidized state only). The external electron donor and acceptor are assumed to be in excess concentration, i. e. the fluxes $J_{10,\mathrm{in}}$ and $J_{5,\mathrm{out}}$ into the protein entrance site (heme 10) and out of the protein exit site (e. g. heme 5) are not limited by donor/acceptor availability and hence do not depend on P_{donor} or $P_{acceptor}$:

$$J_{10,\text{in}} = k_{10,\text{in}}(1 - P_{10}) - k_{10,\text{out}}P_{10} \tag{4.2}$$

$$J_{5,\text{out}} = k_{5,\text{out}}P_5 - k_{5,\text{in}}(1 - P_5) \tag{4.3}$$

Requiring steady state, i.e. $J_{10,\text{in}} = J_{ji} = J_{5,\text{out}} = \text{const} \forall\, i$ gives the following recursive relationship for the steady-state populations:

$$P_{i+1} = [k_{i+1\,i}P_i - k_{10,\text{in}}(1 - P_{10}) + k_{10,\text{out}}P_{10}]/[k_{i\,i+1} + P_i(k_{i+1\,i} - k_{i\,i+1})] \tag{4.4}$$

which can be solved for all P_i. Inserting P_i and P_j in the above expression for J_{ji} then yields the steady state flux $J = J_{ji}$.

A MATLAB [114] script was written to solve these equations for a given set of rate constants $k_{i+1\,i}, k_{i\,i+1}, k_{10,\text{in}}, k_{10,\text{out}}, k_{5,\text{out}}, k_{5,\text{in}}$. In the following, several applications of this formalism are presented and discussed.

4.1 Modelling of physiological conditions

4.1.1 Modelling of electron flux through MtrF

Our model of steady-state electron flux through MtrF is motivated by an experimental study [10] which measured electron flux through the transmembrane MtrCAB complex embedded in a liposome. The otherwise insulating lipid vesicle was filled on the inside with a strong reducing agent (methyl viologen, MV) and oxidation of the dye was measured as soon as an electron acceptor (in the form of some insoluble Fe(III) oxide) was added to the solution (thus outside the vesicle). This yielded different overall oxidation (and thus electron flux) rates depending on the acceptor added, implying that the final ET step from heme to oxide is at least a partial bottleneck.

Here, we model electron flux through MtrF with boundary conditions based on this experimental setup and compare the outcome to the experimental flux rates of White et al. for their MtrCAB complex. We assume that reduction by MV is faster than any intraprotein ET rate (owing to the low redox potential of MV [10]) so that the heterogeneous reduction of the first heme is ultrafast and irreversible. We also assume that the oxidation of the last heme by the final electron acceptor is irreversible but we do not make any assumptions on the heterogeneous ET rate; instead, we treat this last heterogeneous step k_{out} as a variable and calculate the steady-state flux J as a function of k_{out}. In the limit $k_{\text{out}} \to \infty$, J should then converge to an intrinsic maximal flux. We consider heme 10 as the electron entrance site (in accordance with the suggestion in [2]) and calculate J for electron transport to heme 5, 2 or 7

(i. e., to each of the three potential egress sites).

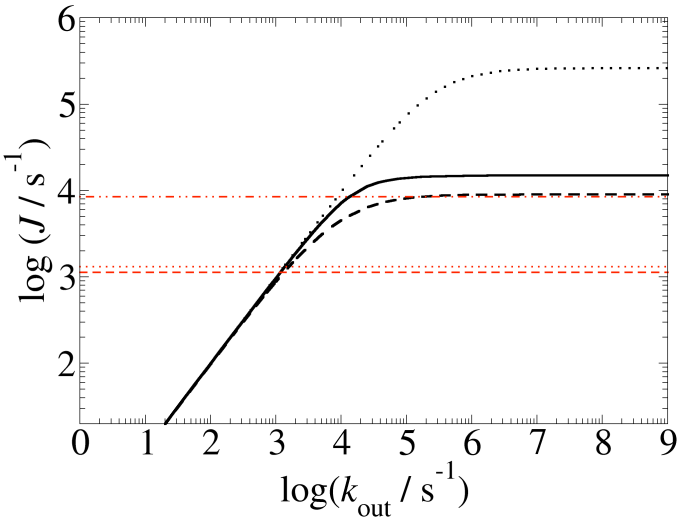

Figure 4.1: Dependence of the steady-state electron flux through MtrF, J (from Eqs. 4.1-4.4), on the heterogeneous rate constant to an external electron acceptor, k_{out} (see text). The electron entrance site is heme 10 and the electron exit site is heme 5 (black solid line), heme 2 (black dashed line) and heme 7 (black dotted line); the respective maximal intrinsic electron fluxes (for $k_{\text{out}} \to \infty$) are 1.5×10^4, 0.9×10^4 and $2.6 \times 10^5 \text{s}^{-1}$, respectively. (See Fig. 1.6 B for the heme connectivity network.) Experimentally measured electron fluxes through MtrCAB to solid Fe(III)-oxide particles are shown for comparison as red horizontal lines for lepidocrocite (dash dotted), hematite (dotted) and goethite (dashed)[10]. Figure reprinted from Ref. [4].

The results for J as a function of k_{out} are presented in Fig. 4.1. We find that for small output rates J increases linearly with k_{out} to indeed asymptotically reach a maximum flux for large values of k_{out}.

In fact, the dependence of J on k_{out} can be expressed in the form $J = J_{\max}/(1 + J_{\max}/k_{\text{out}})$, as is shown in Fig. 4.2 for heme 5 as exit heme. (As well as the opposite direction for comparison, i. e. heme 5 $\to$ 10.) This equation is obtained by mapping the multiple-step electron transport through the protein onto a single step, thereby reducing the overall electron transport from donor to acceptor to an effective four state model based on four sites only: Site A corresponds to an external donor injecting electrons into the protein; site B denotes the protein electron entrance site; site C represents the protein electron egress site; and site D denotes an external electron acceptor. The kinetics is then described by the following scheme:

$$\text{A} \xrightarrow{k_{\text{in}}} \text{B} \underset{k_{\text{b}}}{\overset{k_{\text{f}}}{\rightleftharpoons}} \text{C} \xrightarrow{k_{\text{out}}} \text{D} \tag{4.5}$$

where electron injection into the protein (A to B) and ejection from the protein (C to D) is again considered irreversible. As mentioned above, the compli-

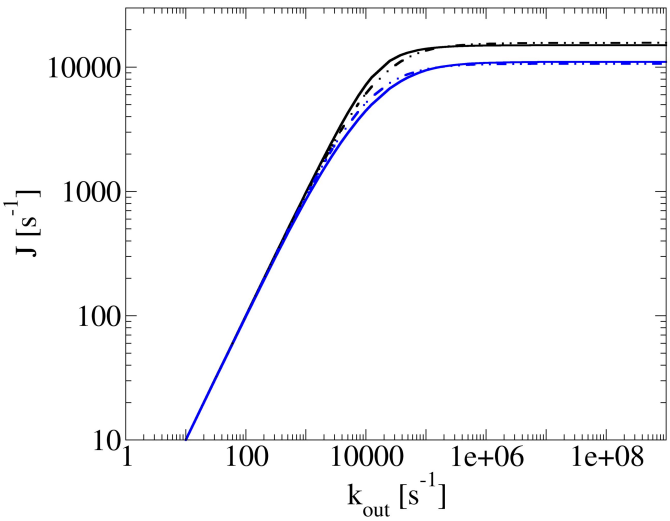

Figure 4.2: J vs. k_{out} (from Eqs. 4.1-4.4, as in Fig. 4.1) for electron flux along the octa-heme chain in MtrF (see Fig. 1.6 B) in both directions (solid lines; black: forward/10 $\rightarrow$ 5, blue: backward/5 $\rightarrow$ 10), together with analytic fits to the numeric curves using Equation 4.10 (broken lines). Figure reprinted from Ref. [4].

cated kinetics of electron flux through the protein is condensed into one effective forward and backward rate constant k_{f} and k_{b}, respectively. Steady-state conditions then require:

$$J = k_{\text{in}}(1-P_{\text{B}}) = k_{\text{f}}P_{\text{B}}(1-P_{\text{C}}) - k_{\text{b}}(1-P_{\text{B}})P_{\text{C}} = k_{\text{out}}P_{\text{C}} \tag{4.6}$$

where P_{B} and P_{C} denote populations in the interval $[0,1]$. Solving for P_{B} in terms of P_{C} and reinserting to solve for P_{C} yields the equation:

$$P_{\text{C}} = \left(\frac{k_{\text{f}}}{k_{\text{in}}} - \frac{k_{\text{b}}}{k_{\text{in}}}\right)P_{\text{C}}^2 - \left(\frac{k_{\text{f}}}{k_{\text{in}}} + \frac{k_{\text{f}}}{k_{\text{out}}}\right)P_{\text{C}} + \frac{k_{\text{f}}}{k_{\text{out}}} \tag{4.7}$$

The assumption of electron injection being much faster than any intraprotein ET yields $k_{\text{in}} >> k_{\text{f}}, k_{\text{b}}$ and Eq. 4.7 reduces to the simple expression

$$P_{\text{C}} = \frac{k_{\text{f}}}{k_{\text{f}} + k_{\text{out}}}. \tag{4.8}$$

Insertion of Eq. 4.8 in the last identity of Eq. 4.6 gives

$$J = \frac{k_{\text{f}}}{1 + \frac{k_{\text{f}}}{k_{\text{out}}}}. \tag{4.9}$$

Thus, for high injection rates the flux through the protein depends only on the effective forward rate k_{f} and the ejection rate k_{out}. Furthermore, if we take the limit $k_{\text{out}} \rightarrow \infty$, J becomes equal to k_{f}: The latter in fact represents the maximal possible flux J_{max} through the protein. We can therefore rewrite

Equation 4.9 as:

$$J = \frac{J_{\max}}{1 + \frac{J_{\max}}{k_{\text{out}}}} \tag{4.10}$$

Using this equation to fit the J-curves for electrons exiting at heme 2, 5 and 7, we obtain $J_{\max} = 0.9 \times 10^4$, 1.5×10^4 and $2.6 \times 10^5 \text{s}^{-1}$, respectively. The corresponding rate-limiting single steps for these routes are 1.2×10^4 (at $1 \rightarrow 2$), 2.9×10^4 (at $1 \rightarrow 3$) and 3.0×10^5 s^{-1} (at $6 \rightarrow 7$). Hence, the order of magnitude difference in the respective rate-limiting steps translates into a corresponding difference in $J_{\max}$. The flow out of heme 7 is particularly large as it avoids one the slow ET steps between coplanar hemes, $6 \rightarrow 1$ ($k_{16} = 4.0 \times 10^4$ s^{-1}).

4.1.2 Modelling of electron flow through MtrCAB

In the following we try to model electron flux through the protein complex MtrCAB, i. e. the transmembrane complex that was used in the study by White et al. [10] MtrA is known to be a deca-heme cytochrome [115] while MtrB is a membrane pore protein proposed to enable close contact between MtrA and MtrC [20]. We thus need a model and ET parameters for MtrC, MtrA and the contact between them. For MtrC, we use our ET parameters for MtrF: As discussed in Section 3.2.3, we assume that our couplings for MtrF might be to a good degree transferable to MtrC, and MtrC might also feature a qualitatively similar free energy landscape. We again assume electron transport along the octa-heme chain from heme 10 to heme 5 (egress site). The structure of MtrA is not known; however, the heme-binding motifs in its N-terminal half can be sequence-aligned with the penta-heme cytochrome NrfB [116], and MtrA has been found to be of a rod-like shape of around 100 Å length [117]. We therefore decided to model MtrA as a NrfB head-to-tail homodimer. We can then use our regressions for the two coupling regimes in Fig. 3.6 to estimate approximate couplings for the heme pairs in the crystal structure of NrfB [118] as well as for the contact between the two NrfB subunits on the one hand and between NrfB and MtrC/F on the other hand (by some crude manual docking of protein structures which should suffice for this modeling). In regard to reorganization free energies, the most significant difference should occur between heme pairs located in the solvent-exposed part of MtrA and those located within the membrane-buried part making contact to MtrC. With a membrane thickness of around 40-50 Å [10], MtrA should be roughly half-buried into the membrane, with the other half exposed into the periplasm. Thus, for the first five heme pairs (as well as the final MtrA-MtrC contact) we assume a reorganization free energy λ of 0.9 eV (i. e. a typical

number for MtrF) while for the four remaining membrane-buried pairs we assume 0.57 eV, the reorganization energy previously obtained [9] for heme *a* to heme a_3 ET in membrane-embedded cytochrome c oxidase. The final set of parameters to be estimated are then the driving forces of each ET step. Whilst the overall electrochemical response of MtrA has been studied [20], redox potentials of individual cofactors are not known. A crude fit to the voltammogram in Hartshorne et al. [20] yields a set of ten distinct redox potentials however (courtesy of Dr. Julea Butt) that enable us to estimate minimal and maximal flux through MtrCAB within the model described so far, by assigning redox potentials to the ten cofactors of MtrA so as to either yield the smallest possible or highest possible rate-limiting single ET rate. We thereby obtain the two curves in Fig. 4.3, delimiting upper and lower limits for the flux through MtrCAB based on our model. As can be seen, depending on the combination of parameters the flux through MtrCAB could reach the same level as for MtrF itself (maximal flux for flux-maximizing parameters, black curve: 14300 s^{-1}); it could also be one order of magnitude smaller than for MtrF (maximal flux for flux-minimizing parameters, blue curve: 800 s^{-1}), but this is rather unlikely as it requires the steepest possible free energy uphill step to coincide with a small electronic coupling and a high reorganization energy.

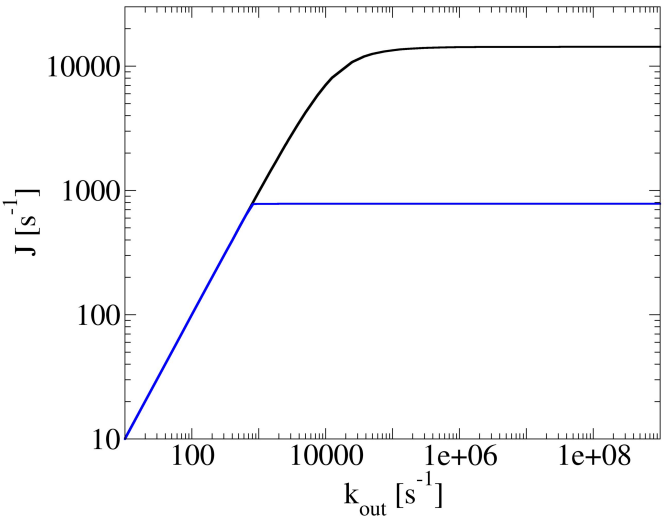

Figure 4.3: J vs. k_{out} (from Eqs. 4.1-4.4, as in Fig. 4.1) for electron flux through the protein complex MtrCAB (see text for model applied). Black curve: Redox potentials in MtrA chosen so as to maximize flux; blue curve: redox potentials chosen so as to minimize flux. Figure reprinted from Ref. [4].

4.1.3 Discussion

The intrinsic maximal flux rates $J_{\max}$ through MtrF and our simple model for MtrCAB can be compared to the experimental data from [10]. To this end, Fig. 4.1 also contains the experimental electron flux rates to three different final electron acceptors (red horizontal lines). The overall rate constant reported was dependent on the type of Fe(III) mineral used, ranging from 1133 to 8500 s^{-1}. This indicates that transport kinetics was mainly limited by ET from the terminal heme to the mineral. Hence, the highest value reported (8500 s^{-1}) should be considered as a lower limit to the intrinsic fluxes calculated here for MtrF and MtrCAB. Interestingly, this highest experimental rate is not much smaller than our calculated maximum flux through MtrF, ranging from $10^4 - 10^5$ s^{-1} depending on the heme exit site. A similar intrinsic flux seems feasible for MtrCAB: Assuming that the free energy landscape for ET through MtrA is not unfavourably shaped (see model parameters in previous section), we obtain a steady-state current similar to the one for MtrF alone. The assumptions that heme edge-to-edge distances are not larger in MtrA than in MtrF and that the reorganization free energies λ in the solvent-exposed part of MtrA are also comparable to MtrF seem somewhat safer.

At this point we can discuss the impact of the correlation between free energy landscape and electronic couplings in MtrF in more detail. The steady-state modeling approach enables us to experiment with the heme-to-heme ET rates k_{ET} as well as their constituent quantities (in particular, couplings and driving forces). If we were to keep the free energy landscape fixed and permute the heme dimer motifs such that the stacked hemes become coplanar and vice versa, the big energetic uphill step from heme 10 to 9 would not be counteracted by a favourable coupling anymore and hence the maximum electron flux along the octa-heme chain would decrease 17-fold from 1.5×10^4 to 9×10^2 s^{-1}. (The single rate for $6 \rightarrow 1$ would of course significantly increase, but this would not preserve the overall flux.) Similarly, keeping the couplings fixed and changing the redox landscape such that the first 0.2 eV uphill step moves from $10 \rightarrow 9$ (stacked) to $6 \rightarrow 1$ (coplanar) would yield a 25-fold decrease in maximal flux to 6×10^2 s^{-1}. While it is not clear what the impact of such a decrease on respiration rates and metabolism would be, it would let $J_{\max}$ drop below the experimentally established lower limit of ~10^4 s^{-1} by at least an order of magnitude. This suggests that the proper matching of couplings and driving forces is indeed necessary to maintain the proper function of MtrF for electron transduction.

4.2 Modeling of a tunneling spectroscopy experiment

4.2.1 Modeling of current-voltage response

Pioneering tunneling spectroscopy measurements employing an AFM tip and a gold electrode revealed that bacterial nanowires can support very high currents of several nano-Ampere (nA) at moderate voltages [32, 31]. It has been assumed that the conduction along these nanowires is facilitated by multi-heme cytochromes and that conduction through multi-heme cytochromes in these setups occurs via electron hopping along the heme groups [111, 113]. These nanowires have recently been shown to be outer membrane extensions, indeed containing multi-heme cytochromes along their entire length [34]. In addition, recent work has expanded these measurements to single MtrF molecules immobilized on a surface [48], yielding nA currents even for a single molecule.

Here, we would like to investigate if the hopping mechanism can account for the observed nA currents when we apply the heme-to-heme ET parameters obtained in this work. To this end, we model the current-voltage response of a single MtrF protein placed between two electrodes with potential difference V by the same steady-state approach as used in Section 4.1; however, now we cannot use our parameters anymore without modification as the external bias potential will affect the heme-to-heme driving force. The experimental setup for tunneling spectroscopy on a single MtrF molecule partially preserved the hydration environment, but it is not clear how much water still remains around the protein even when the sample is macroscopically "dry" [48]; here, we consider two limiting cases, complete solvation on the one hand (where reorganization free energies and intrinsic driving forces can be taken from our simulations in solvated environment) and a dry protein on the other hand (requiring adjustments in λ and ΔA). As mentioned above, we would like to study whether the hopping model is consistent with the experimentally observed nA currents; thus, the aim is not a quantitative estimate but an order of magnitude estimation.

4.2.2 Model details

We obtain the current I for a given bias potential V by the convention $I = -J$. For calculation of the heme-to-heme ET rates in solution conditions, we use the values for H_{ab} and λ that we calculated in Chapter 3; the driving forces from Chapter 3 are modified so as to incorporate the external bias potential: $\Delta A \rightarrow \Delta A - eV/(n+1)$ where we assume that the potential falls off linearly

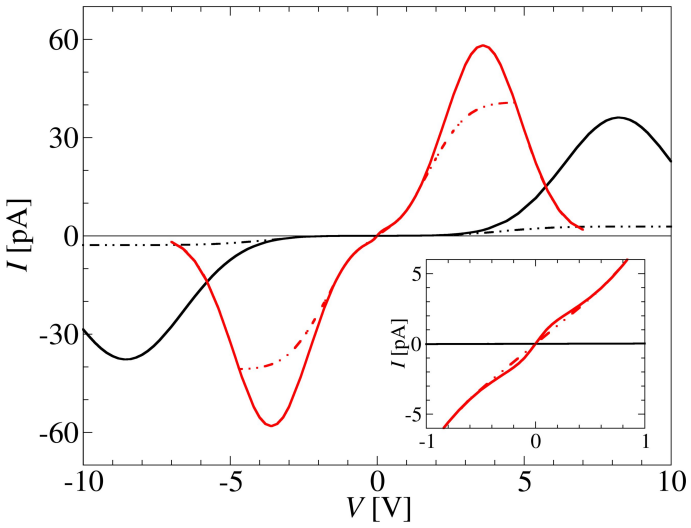

Figure 4.4: Modelled Current (I)-Voltage (V) response of a single MtrF molecule in solution (black lines) and in air (red lines) assuming heme-to-heme electron hopping as the conduction channel. (Calculated via Eqs. 4.1-4.4, but modifying driving forces via a bias voltage V; see text.) Two different regimes are shown, protein limiting ($r = 100$ (black solid line), $r = 10$ (red solid line)) and electrode-protein limiting ($r = 1$ (black and red dash dotted lines)). See text for definition of r. Figure reprinted from Ref. [4].

and the cofactors are spaced roughly evenly between the two electrodes. For lack of more specific information, we assume electron transduction along the octa-heme chain between hemes 5 and 10. The heterogeneous ET steps between heme 10 and the left electrode and heme 5 and the right electrode are treated as reversible (i. e. $k_{10,\mathrm{out}}, k_{5,\mathrm{in}} \neq 0$) and symmetric (i. e. $k_{10,\mathrm{in}} = k_{5,\mathrm{out}}$ and $k_{10,\mathrm{out}} = k_{5,\mathrm{in}}$) and are obtained via the electrochemical form of the nonadiabatic ET rate equation [119, 111]:

$$k_{10,\mathrm{in}} = C_{\mathrm{elec}} \int_{-\infty}^{+\infty} \frac{\exp\left[-\left(x - \frac{\lambda + e(E-E_{10})}{k_{\mathrm{B}}T}\right)^2 \left(\frac{k_{\mathrm{B}}T}{4\lambda}\right)\right]}{1+\exp(x)} dx \tag{4.11}$$

$$k_{10,\mathrm{out}} = C_{\mathrm{elec}} \int_{-\infty}^{+\infty} \frac{\exp\left[-\left(x - \frac{\lambda - e(E-E_{10})}{k_{\mathrm{B}}T}\right)^2 \left(\frac{k_{\mathrm{B}}T}{4\lambda}\right)\right]}{1+\exp(x)} dx \tag{4.12}$$

where λ is the reorganization energy of the heterogeneous ET step, E is the potential level at the electrode, E_{10} is the potential at heme 10, C_{elec} is a constant denoting the average coupling between heme and electrode and k_{B} and T are Boltzmann constant and temperature, respectively. Hence, to estimate the heterogeneous ET rates at the electrodes via this pair of equations, estimates are needed for the reorganization free energy λ, the local potential drop $E - E_{10}$ and the coupling constant C_{elec}. Whilst no values are available

for the reorganization free energy of MtrF in contact with an electrode, electrochemical measurements of wild type and mutated cytochrome c on a gold electrode [120] yielded reorganization free energies of around 0.45 eV which we chose to use in our model of current-voltage response in solution. For the potential drop at the electrode we made the assumption that it is comparable to the voltage drop between adjacent hemes, i. e. $1/(n+1)=0.11$ (with eight redox sites); this is similar to experimental estimates [32]. For the electrode contacts, $1/2\times\Delta E^0_{10\to 5}$, half the redox potential difference between the terminal hemes 10 and 5, was added to correct for the unequal potential levels of the terminal hemes at zero bias potential (yielding a nonzero current at zero potential otherwise).

The remaining parameter is then C_{elec}, the constant summarizing the overall coupling between terminal redox site and electrode. Rather than just making one assumption for this electrode-protein coupling, we tried different values to obtain the current-voltage response in different regimes. Specifically, we determined C_{elec} by fixing a certain ratio r of heterogeneous input (= output) rate, $k_{10,\text{in}}(=k_{5,\text{out}})$ and the smallest heme-heme ET rate, $k_{ji}^{\min}$ evaluated at zero potential bias, $r=k_{10,\text{in}}/k_{ji}^{\min}$. A high r (10-100) yields protein-limited current while a small r of 1 leads to electrode contact limitation for certain voltage ranges.

Under dry conditions, i. e. for a measurement in ambient air, we assume the couplings to stay the same given that these depend only on the heme cofactor arrangement which should change the least. In regard to the reorganization free energies, a significant change can be expected given that they are known to be highly affected by the solvent environment. Specifically, Tipmanee et al. [71] found for a set of model ET proteins that the solvent contributed $1/2$ to $3/4$ of the total reorganization free energy. Without more precise information on the corresponding solvent contributions in MtrF, we chose a λ of 0.4 eV for intraprotein ET and 0.2 eV at the protein-electrode interface for our ambient air model. In regard to driving forces, no information is available on these under dry conditions, but it can be assumed that without a solvation shell, ionizable groups should be either neutral or binding a counter ion so that the electrostatic potential on air should be much more homogeneous than in water. We therefore decided to set all intrinsic driving forces equal to zero so that the total driving forces are equal to the contribution from the external potential. Obviously this crude set of parameters does not allow for detailed predictions and is only used to get an idea what the current-voltage behavior might be under dry conditions.

4.2.3 Results and Discussion

In Fig. 4.4 the simulated *I*-*V*-curves obtained for different assumed conditions are shown. Curves in black assume solution conditions, curves in red assume a dry protein (air). Solid and dash dotted lines represent the two different regimes mentioned above: Protein limited ($r = 100$, black solid line; and $r = 10$, red solid line) and partially electrode limited (black and red dash dotted lines). For $r = 100$ (solution) and $r = 10$ (air), respectively, the protein limitation of total flux gives rise to a clear maximum in I in Fig. 4.4 (solid lines) indicating the transition between normal and inverted Marcus regime for nonadiabatic ET. Decreasing r to 1 yields the broken lines which for dry conditions at first show an electrode-limited response, reaching a constant current indicating the maximal overlap between Gaussian redox peak and Fermi distribution in Eq. 4.11 - until the slowest protein rate becomes slower than the heterogeneous rate upon which the curve matches the corresponding curve for higher r again. For solution conditions, only the electrode-limited regime is visible in the potential range studied.

The maximum current we obtain under solution conditions is 36 pA at a voltage of 8.2 V; when we assume dry conditions (i. e., decreasing λ and removing intrinsic driving forces), the maximum current becomes 58 pA at 3.6 V. For comparison, the single-molecule measurements of Byun et al. [48] yield currents of 1-3 nA (depending on the humidity conditions) for a bias voltage of 2 V (highest voltage shown in their results). Thus, our calculations suggest that the surrounding medium (air/solution) has a large effect on the voltage range but only a relatively small effect on the maximum current, which in either case is smaller than in experiment by a factor of 30-50 with our maximum currents or 100 if we use our "dry-protein" current at 2 V. (The "wet protein" current is still virtually zero at that bias voltage.) Hence, if the main effect of removing the solvation environment is just to lower λ while not changing the mechanism of ET, then the different conditions in experiment cannot account for the almost two orders of magnitude difference between the experimental currents in the nA range [32, 31, 48] and the tens of pA we obtain with our model. (If (partial) drying of the protein were to, say, increase the couplings somehow, the different conditions could perhaps account for the difference.) Hence, while the heme-to-heme hopping with full thermal equilibration after each hop is able to reproduce the experimental data on electron transduction through MtrCAB in solution (see Section 4.1), it seems quite possible that the mechanism of ET in tunneling spectroscopy experiments is a different one altogether.

Chapter 5

Substrate interaction in MtrC

In this last part of our work we turn to the interaction of OM cytochromes with soluble substrates. As discussed in Chapter 1, it has been established experimentally that OM cytochromes can reduce small redox active molecules, specifically flavins, and these substrates have been implied to act as electron shuttles between the cytochromes and solid final electron acceptors, hence enabling reduction of the insoluble substrate without the need for direct cytochrome-surface contact. However, the main open questions are where on the cytochromes the flavins bind and how strongly. The former question not only concerns the specific binding site for a flavin but more generally whether such a site exists to begin with, or whether the flavin binds unspecifically to the cytochrome. The latter question concerns the qualitative mode of interaction (transient interaction vs. stable binding as a cofactor), and as outlined further in Chapter 1, very recent experimental evidence [3] suggests that the mode of interaction for some cytochromes actually depends on the redox state of a certain disulphide bond, with flavins interacting transiently with the protein with the disulphide bond oxidized/closed and binding stably if the bond is reduced/open.

In the following, we describe our computational approaches to these questions mainly based on empirical docking studies. We focus on the binding of FMN to MtrC as flavin interaction with this cytochrome has been of more interest in the field [121, 28, 35], and the effect of cleaving the disulphide bond on the FMN affinity is also reported to be more pronounced in MtrC than in MtrF. [3] We study in molecular detail the docking of FMN to MtrC in the state with the disulphide bond oxidized/closed (henceforth referred to as the "SS" state from the disulphide “-S-S-” bond) and afterwards carry out Simulated Annealing studies to investigate possible conformational changes in MtrC upon cleaving the disulphide bond (yielding the “SH” state, from the reduced thiol groups) that could explain the observed steep increase in binding affinity for FMN.

The circumstances for the docking studies are somewhat special in that no *a priori* information is available on possible binding sites (not even if there are well-defined binding sites to begin with). To investigate flavin binding under these circumstances we therefore turn to a "blind docking" protocol which makes no assumptions on the binding site for a ligand. The empirical docking program Autodock [86, 87] enables such a blind search as it carries out an unbiased search on a predefined region of space around the target protein (see theory in Section 2.4). In fact, such a blind docking protocol has already been employed and tested for ligands of comparable size to FMN and found to be effective in reproducing known binding sites [122, 123]. However, FMN itself was not amongst the ligands tested and hence the first part of this project involved setting up and validating such a blind docking protocol for FMN by re-docking FMN for experimentally known protein-FMN complex structures.

5.1 Procedure

5.1.1 Validation of the FMN blind docking protocol

We chose two FMN-bound protein structures to test and validate our blind docking protocol for FMN: The flavin-binding protein (FMN-bp) from *Desulfovibrio vulgaris* (Miyazaki F) (pdb code 1AXJ) [12] and the NAD(P)H:acceptor Oxidoreductase (FerB) from *Paracoccus denitrificans* (pdb code 3U7R) [11]. For both of these proteins there are also experimental binding affinities available, allowing to test not only the prediction of binding poses but also of binding free energies. The NMR structure for FMN-bp contains 20 configurations; configuration 14 was chosen for the re-docking. In order to re-dock, the bound FMN molecule was removed from both proteins; in the case of FerB, crystal water and one nonaethylene glycol were stripped as well and one selenomethionine was mutated to a regular methionine as no parameters were available for selenium. Furthermore, protonation states for ionizable groups in FerB were set according to the results of the pKa estimator propKa 3.1 [93]. For FMN-bp, the NMR structure already included hydrogens. We used the two protein structures as-is without force field relaxation/minimization and merged nonpolar hydrogens in both FMN and proteins to their parent carbons using the utility program AutodockTools. (This means that the hydrogens are deleted and any atomic charges on them added to the carbon bound to them.)

Autodock 4.2 [87] was used for all docking calculations. The default atomic parameters in Autodock were used except for the atomic charges which where taken from the AMBER03 force field. [89] Atomic charges for FMN were obtained according to the RESP procedure [124] from a DFT electronic structure calculation (using NWChem [103]) with the B3LYP exchange corre-

lation functional [78, 79] and cc-pVTZ basis set, combined with the COSMO continuum solvation model [125] (using a relative permittivity of 4.0). The atomic charges of FMN are summarized in Table 5.1. (RESP parametrization courtesy of Dr. Adam Kubas.) During docking, intramolecular electrostatic interactions and hydrogen bonding terms between atoms of FMN were switched off. This was found to be necessary to prevent spurious formation of intramolecular hydrogen bonds between the ribitol hydroxy groups and the phosphate group of FMN. As we will see in Section 5.2.1, the re-docking of FMN to the two flavin-binding proteins with the same intramolecular interactions switched off were rather successful, justifying this ad-hoc approach.

Table 5.1: RESP charge parametrization for FMN.

Atom name	Charge [e]	Atom name	Charge [e]
N1	-0.396715	C5'	0.069370
C2	0.373383	O5'	-0.411130
O2	-0.560608	P	0.923193
N3	-0.040698	O1P	-0.763397
C4	0.149203	O2P	-0.763397
O4	-0.517263	O3P	-0.763397
C4A	0.590519	H3	0.237377
N5	-0.544057	H6	0.125184
C5A	0.162506	H7M1	0.083170
C6	-0.134921	H7M2	0.083170
C7	0.029809	H7M3	0.083170
C7M	-0.240263	H8M1	0.047064
C8	0.069789	H8M2	0.047064
C8M	-0.081776	H8M3	0.047064
C9	-0.180532	H9	0.116942
C9A	0.016667	H1'1	0.037677
N10	0.076287	H1'2	0.037677
C10	0.050082	H2'	0.098083
C1'	0.091311	HO2	0.357665
C2'	-0.022327	H3'	0.074538
O2'	-0.542320	HO3	0.253008
C3'	0.015111	H4'	0.115456
O3'	-0.366671	HO4	0.380892
C4'	0.051154	H5'1	0.031974
O4'	-0.628058	H5'2	0.031974

The blind docking was carried out employing the Lamarckian Genetic Algorithm as implemented in Autodock. This is a modified genetic algorithm featuring occasional local optimizations of individuals. All parameters relating to the genetic algorithm were kept at their default values except mentioned otherwise. A $23 \times 38 \times 38$ Å^3 search box was used for FMN-bp and a $35 \times 37 \times 27$ Å^3 search box was used for FerB, as indicated in Figure 5.3

(A) and (B). These boxes were chosen to resemble the search conditions for MtrC (see further below) and were large enough to cover most of the protein surface of FMN-bp and a large surface on the bigger protein FerB, hence allowing for the flavin to probe protein regions far away from the experimentally determined binding site. 1200 docking runs with 75 individuals and 7000 generations per run were carried out for FerB and 300 docking runs with 100 individuals and 7000 generations per run for FMN-bp. The resultant 1200 docking poses for FerB were clustered with an RMSD cutoff of 3.0 Å and the 300 poses for FMN-bp with a cutoff of 2.0 Å. This means that all poses with an RMSD relative to the global lowest-energy pose that is smaller than the cutoff were included in the first cluster. The lowest-energy pose among the remaining poses was then the reference for the second cluster and so on. This yielded large numbers of clusters in total (around 50 for FMN-bp and almost 200 for FerB), of which only the lowest-free energy ones were significantly populated however.

5.1.2 Docking of FMN to MtrC in the SS state

Our docking studies to MtrC in the SS state started off from the crystal structure [3]. The crystal structure contains five Ca ions which might however originate from the crystallization buffer; we removed four of the ions and left one (around heme 3) that appeared to be a little more buried in the structure. We set all hemes to be reduced assuming that this should be the physiologically more relevant state if MtrC was to reduce a docked flavin. Protonation states were chosen according to the results of the pKa estimator propKa 3.1 [93]. The crystal water was removed. The protein structure was then relaxed (energy-minimized) in NAMD 2.9 [95] using the AMBER03 force field [89] with heme parameters as in our previous studies [5, 71, 9]. The relaxed structure was used for the docking runs. As for the two proteins from the redockings, nonpolar hydrogens were merged in AutodockTools.

The blind docking was carried out similarly as for the FMN-binding proteins above. The same default Autodock atomic parameters and the same atomic charges for FMN were used as before, with atomic charges for protein atoms taken from the AMBER03 force field [89] and the ones for the heme cofactors and axial histidine ligands from the same parametrization we used for MtrF as well [5, 71, 9]. As in MtrF, eight of the ten hemes in MtrC are solvent accessible with hemes 3 and 8 buried inside the protein; the latter two were thus not considered as viable docking targets. We therefore carried out an individual blind docking procedure for each of the other eight hemes, 10, 9 , 7, 6, 1, 2, 4, and 5. For each heme to which FMN was docked, a search box was centered on the heme with enough space in all directions to allow

docking in the heme's extended surrounding. This yielded box lengths in the range of 25 to 40 Å in each direction. For each heme 1200 individual docking runs were carried out. The populations for the genetic algorithm were chosen between 75 and 125 individuals for each run depending on the box size. Each docking run consisted of 7000 generations in the genetic algorithm. The resultant 1200 poses were clustered with an RMSD cutoff of 3.0 Å resulting in about 100 clusters for each heme.

5.1.3 Simulated annealing in the SH state

Observation from experiment suggests that upon reduction of the disulphide bond, MtrC undergoes some conformational change that leads to a significant increase in the binding affinity of FMN. We tried some exploratory runs at room temperature or slightly elevated temperature starting from an SS equilibrated structure with the disulphide bond cleaved, but these showed no significant changes in the structure; therefore, we turned to Simulated Annealing (SA) in order to accelerate the conformational sampling by heating up the protein to higher temperatures and subsequently cooling it down to room temperature again. Prior to SA, the protein was equilibrated at room temperature. For this part of the study, all hemes were set to oxidized corresponding to the conditions in the study of Edwards *et al.* [3]; the same protonation states as in the docking runs were used but all of the five calcium ions were included (as a non-structural ion could now simply diffuse away during the dynamics). The protein was solvated with a water layer of thickness 15 Å and sodium and chloride ions were added to neutralize the system and obtain a salt concentration of about 0.1 M. The disulphide bond was treated as closed (SS) initially. All MD simulations were carried out with NAMD 2.9 [95] in periodic boundary conditions. The system was energy-minimized for 5000 steps before the solvent was equilibrated for 500 ps with the protein kept frozen, using periodic temperature rescaling to 300 K and a barostat to equilibrate the volume. Keeping the temperature rescaling but fixing the volume, the protein was then slowly released by restraining it with successively weaker harmonic force constants of 99, 75, 50, 25, 10, 5, 1, 0.1 and 0.01 kcal/mol/Å^2 (each step for 250 ps with a 1 fs time step until 5 kcal/mol/Å^2 and for 500 ps with a 2 fs time step thereafter). All restraints were then fully released, the barostat switched on again and the thermostat changed to a Langevin thermostat. The system was equilibrated for 7 ns after which the disulphide bond was cleaved and the two sulphur atoms were saturated with hydrogen. In this SH state the system was equilibrated for another 5 ns. The output of this last equilibration step served as input for the SA studies.

Test simulations at elevated temperature revealed that Domain III is sur-

prisingly stable while the other domains, particularly Domain I, were less stable. It was found to be necessary to restrain certain protein regions during SA to avoid denaturation of the protein structure at high temperature. In order to exert restraints as mild as possible, the targeted MD feature in NAMD was used. This option allows one to restrain the total RMSD of a specified protein region with respect to a reference structure, rather than restraining atoms to individual reference positions. Two separate restraint regions were defined: one region comprised of the backbone atoms of the entire Domain I, parts of Domains II and IV and hemes 1, 2, 8, 9 and 10; the other region containing the alpha-helix connecting Domains II and III. These regions are depicted in Fig. 5.1. They were chosen so as to keep the restraints clear from Domain III as far as possible. A force constant of 75 kcal/mol/Å^2 was used to harmonically restrain each region to zero RMSD with respect to its initial structure for the SA runs, which is the final structure of the equilibration at room temperature (see above).

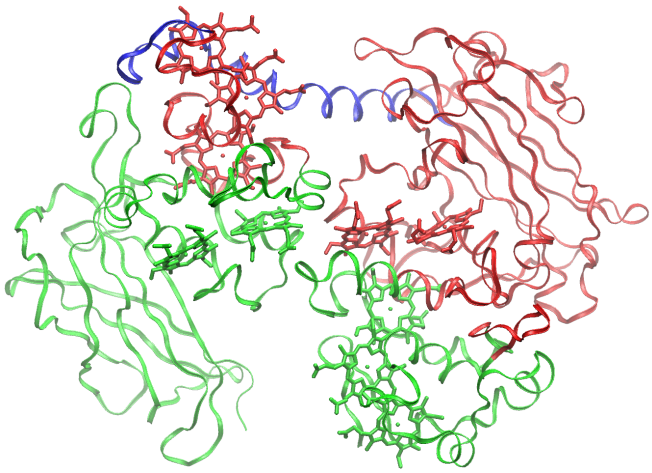

Figure 5.1: Setup for RMSD-restrained Simulated Annealing (SA) runs: The two restraint regions are depicted in green and blue, respectively. These were separately restrained to a target RMSD of 0 Å with respect to their initial structure (i. e., the last snapshot of the preceding equilibration). The region in red was unrestrained.

The annealing protocol consisted of instantaneously heating the system to a high temperature (resampling velocities at that temperature and setting the thermostat accordingly) and simulating the system at this temperature for 1 ns. This was then followed by cooling the system down to 300 K by lowering the thermostat temperature every 100 ps in steps of 50 K. At 300 K, the simulation was run for 1 ns with the RMSD restraints active and then for another 1 ns with the restraints turned off. The initial temperatures were determined by investigating the protein behaviour over a range of temperatures from 500 K to 900 K: While no potentially significant conformational changes were observed for initial temperatures below 600 K, Domain III often showed a de-

naturated structure upon cooling for initial temperatures above 700 K. Hence, initial temperatures of 600 K, 650 K and 700 K were chosen. 24 SA runs with initial temperatures of 600 K and 650 K and 28 SA runs at 700 K starting temperature were carried out. In order to isolate conformational features specific to the SH state, similar SA runs were carried out for the SS state for comparison (20 runs for each of the three temperatures).

Finally, as the recooled structures from the SH annealings did not appear necessarily equilibrated (with several hydrophobic residues quite solvent-exposed), we selected the final snapshot of one 700 K SA run and let it further run at room temperature for a long time (around 110 ns). The final snapshot of one 700 K SA run was selected and further equilibrated for 110 ns at room temperature. As is discussed in detail in Section 5.2.3, this room temperature continuation resulted in a conformational change of the loop containing the two cysteines ("cys-loop" from now on) making a large-scale motion towards Domain II, bringing the two now unbound cysteines around 20 Å apart in the process. After this conformational change took place during the first few 10 ns, it was observed to be stable during the remaining simulation time. Four randomly chosen snapshots after 60 ns run time and the final snapshot at 110 ns were selected for subsequent docking studies (see next section).

5.1.4 Docking of FMN to annealed MtrC in the SH state

Several potential regions of interest arose due to the large scale motion of the cys-loop: The top of the barrel where the loop had moved away; the front of the protein where the loop was making contact to Domain II; the region around heme 4 and 5 where the cys-loop now passed by; and the region around heme 7, which was suggested as a potential binding site in experiment. [3] With the five snapshots chosen for docking, this resulted in 20 docking jobs in total, with box dimensions and genetic algorithm populations similar to the dockings in the SS state. To scan for potential new binding sites more rapidly, we only ran 300 runs for each snapshot and binding region (leaving the clustering RMSD cutoff at the default value of 2.0 Å for these dockings). For the dockings to heme 4 and heme 7 we found one interesting snapshot each for which we carried out more extensive docking runs (1200 runs in total, clustering at 3.0 Å as for the SS state dockings). The hemes were chosen to be in the all-reduced state to facilitate comparison of binding affinities with docking runs carried out for the SS state. Additional dockings to hemes 4 and 7 were also carried out for the all-oxidized state, which was the experimental redox state in the disulphide cleavage experiment [3] and the state for which the SA simulations were carried out. The docking results were found to be rather insensitive to the exact heme redox state for a given protein configuration, with

the main effect a slight increase in affinity.

5.2 Results

5.2.1 Validation of FMN docking protocol

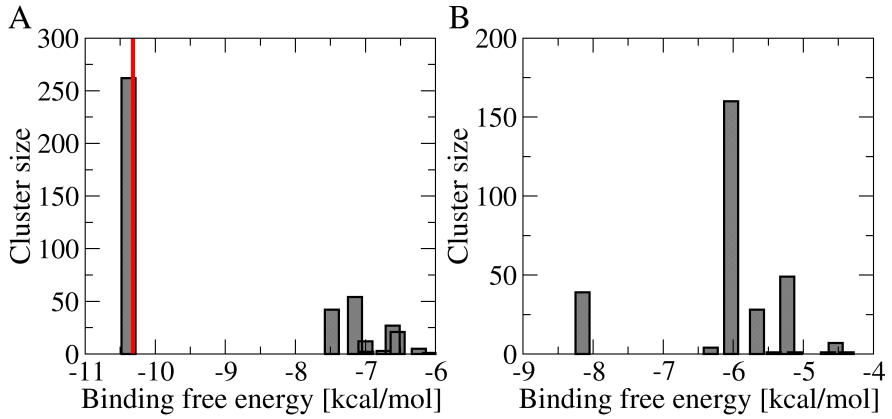

Figure 5.2: Histograms for re-docking to two FMN-binding proteins (only first ten clusters shown). (A) Re-docking to FerB. The experimental binding free energy obtained from the experimental dissociation constant K_D (Ref. [11]) is indicated by a red vertical bar. (B) Redocking to FMN-bp.

The clusters obtained from docking were ordered according to the pose with lowest binding free energy in each cluster. In Fig. 5.2 the size of each cluster (first ten shown) is plotted versus the lowest binding free energy (within that cluster) for both FMN-binding proteins. In both cases the first cluster containing the best pose with the overall lowest binding free energy is separated from the next cluster by a significant energy gap (2-3 kcal/mol). In the case of FerB (Fig. 5.2(A)) the cluster containing the best pose is also the most populated one, with almost one in four individual docking runs (262/1200) resulting in a pose in that cluster. In case of FMN-bp (Fig. 5.2 (B)) there are two other clusters with comparable population to the cluster with the best pose (one of them, the fourth cluster, even being a bit larger). However, these are already 3 kcal/mol higher in energy than the best overall pose, therefore still allowing for a clear discrimination. In addition, the dockings for FerB had been distributed over eight jobs with 150 dockings per job that were individually clustered with the Autodock default cutoff of 2.0 Å. All of these 8

individual clusterings yielded lowest energy poses in close agreement with the best pose obtained by clustering all of the 1200 poses, and the corresponding 8 histograms were similar to the overall histogram shown in Fig. 5.2.

The best re-docked FMN poses with the lowest binding free energy are shown in Fig. 5.3 (A) and (B) for FMN-bp and FerB, respectively (structures in red). They are overlayed on the experimental crystal/NMR structure of FMN, shown in blue. Both experimental structures of the ligand are reproduced very well with RMSD values for the heavy-atoms of FMN of 1.1 Å (FMN-bp) and 1.3 Å (FerB), respectively. Most of the residual deviation arises from the side chain, as the RMSD values for the head group alone amounts to 0.5 Å (FMN-bp) and 0.3 Å (FerB), respectively. Including both head group and phosphate tail increases the RMSD to 1.0 Å for FMN-bp but still yields only 0.4 Å for FerB.

The ligand dissociation constants as obtained from Autodock are $K_{\mathrm{d}} = 1.1\,\mu$M and 25 nM for the best poses for 1AXJ and 3U7R, respectively. The latter value matches the experimental dissociation constant of 27 ± 2 nM very well[11] (indicated by a red bar in Fig. 5.2(A)). In contrast, when we intentionally excluded the experimental binding site from the search region in 1AXJ, the dissociation constant of the best pose increased from 1.1 to 770 μM and the histogram showed a continuum of clusters. Moreover, while all clusters from the original re-docking for 1AXJ (including the experimental binding region) agree in the location and alignment of the head group, side chain and phosphate tail, the poses from the re-docking without the experimental binding region only agree in the location of the most favourable binding site, but not in the alignment of the ligand.

The docking results also allowed us to analyze the binding of FMN to these two proteins (which could then further down be compared to the situation in MtrC). Fig. 5.3 shows the hydrogen bonds formed between the re-docked FMN (best overall pose) and FMN-bp or FerB, respectively (using a cutoff of 2.5 Å for the distance between hetero and hydrogen atoms). It can be seen that the “mold" into which the flavin docks (illustrated in panels C and D) provides in each case 4 hydrogen bonds for the flavin head group as well as a number of hydrogen bonds with the side chain and phosphate tail, yielding in total 11 (FMN-bp) and 13 (FerB) hydrogen bonds, respectively. This is in good agreement with the experimental numbers of 11 and 12 hydrogen bonds for FMN-bp and FerB, respectively.

Thus, our results for re-docking of FMN to two known FMN-binding proteins gave credence to both the parametrization of ligand and protein and the blind docking protocol; with the overall very good agreement with experiment (both in terms of binding poses and affinities) the chosen protocol was

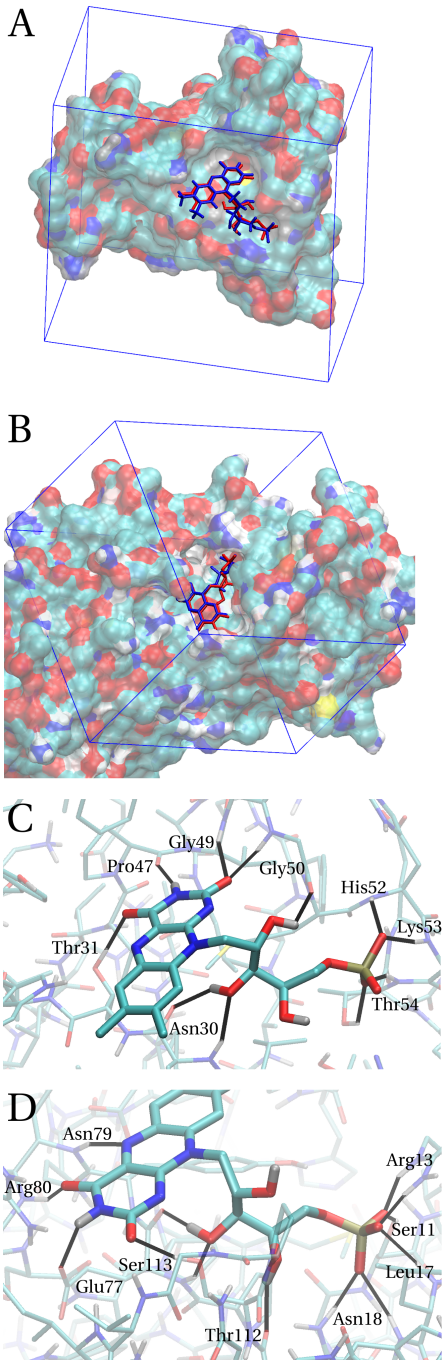

Figure 5.3: Re-docking of FMN to two FMN-binding proteins. (A) Re-docking to FMN-binding protein (FMN-bp) from *Desulfovibrio vulgaris* (Miyazaki F) (pdb code 1AXJ) [12]. The experimental binding pose of FMN is shown in blue, the best pose obtained from computational re-docking is shown in red. The rectangular box indicates the Autodock search region. (B) Re-docking to NAD(P)H:acceptor Oxidoreductase (FerB) from *Paracoccus denitrificans* (pdb code 3U7R) [11]; same color-code as in (A). (C) Close-up of the re-docked pose of FMN shown in (A), indicating individual hydrogen bonds (black) together with the protein residues involved. (D) Close-up of the re-docked pose of FMN shown in (B), indicating individual hydrogen bonds (black) together with the protein residues involved.

deemed useful to predict the interaction of FMN with other proteins such as MtrC.

5.2.2 Docking of FMN to MtrC in the SS state

After preliminary tests showed that searching the entire protein surface at once was ineffective, we chose to search a large region around each heme individually for possible docking sites (see protocol in Section 5.1.2). In fact, recent NMR results [13] indicate that FMN should indeed bind closely to a heme, so that from the onset excluding protein regions further away is justified. Of the eight hemes subjected to docking, six yielded FMN binding poses in (or almost in) van der Waals contact to the heme. Docking to the two central hemes 1 and 6 resulted in final poses that were closer to heme 2 and 7, respectively. Hence, association of FMN with these hemes seems least probable. Table 5.2 summarizes K_d for the best poses for each heme. As can be seen, these range from 490 μM for heme 2 to 30 mM for heme 5.

Table 5.2: Dissociation constants K_d for hemes in MtrC in the SS state.

Heme	K_d [mM]
1	-
2	0.49
3	-
4	12
5	29
6	-
7	7.4
8	-
9	8.9
10	17
protein surface	2.9

For comparison, we also docked FMN to a region of the protein surface in domain I that is far away from any of the hemes, i. e. in a region that is expected to be functionally irrelevant for flavin docking. This yielded a dissociation constant of 2.9 mM for the lowest free energy pose (entry 'protein surface', last row in Table 5.2). Together with the value of 770 μM for the protein surface excluding the binding region in FMN-bp (see results for FMN-bp in Section 5.2.1), this suggests that dissociation constants on the order of 1 mM can easily be achieved on globular protein surfaces that do not contain specific FMN binding motifs. It further suggests that heme 2 is the only heme in MtrC to have an affinity for FMN that is stronger, but not much stronger, than the 'base line' affinity corresponding to 1 mM.

The relative binding affinities also allow for a comparison to the binding stochiometry obtained from NMR measurements [13] which gave FMN:MtrC = 1:1. In our dockings, heme 2 not only has a K_d one order of magnitude lower than the K_d for the heme with the second highest affinity, but it is also the only heme to exceed the aforementioned basis line affinity. Thus, it appears to be reasonable to conclude that the affinity of FMN to heme 2 accounts for a good fraction of the overall affinity of FMN to MtrC. This is further supported by the good agreement between the calculated dissociation constant for heme 2 (490 μM) and the experimental value of 255$\pm$126 μM. [13]

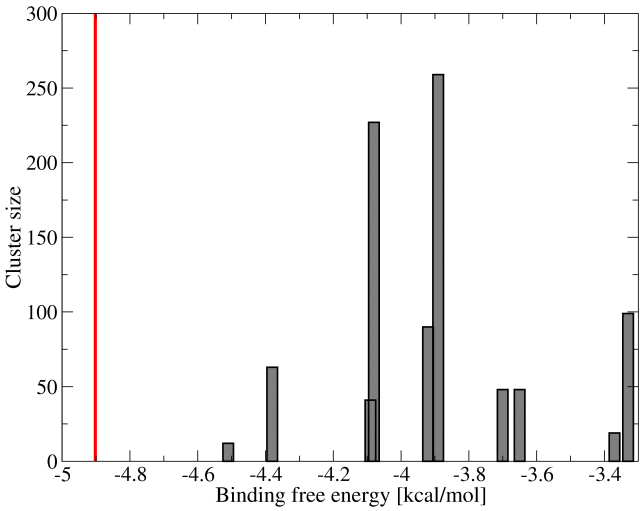

Figure 5.4: Histogram for docking of FMN to heme 2 in the crystal structure of MtrC (only first ten clusters shown). The experimental binding free energy obtained from the experimental dissociation constant K_D (Ref. [13]) is indicated by the red vertical bar.

The structure of the FMN-heme 2 complex is less clearly defined. The histogram for heme 2 in Fig. 5.4 does not show a single most favourable cluster that is significantly lower in energy than the others, in contrast to what was found for re-docking to FMN binding proteins (Fig. 5.2). Rather, the histogram shows an almost continuous spectrum of binding free energies, with the best pose only 0.1 kcal/mol lower in free energy than the second-best one and 0.4 and 0.6 kcal/mol lower than the two largest clusters. The lowest energy poses in the different clusters only agree in the location relative to heme 2 whereas they differ significantly in orientation. This is illustrated in Fig. 5.5A, where the lowest energy poses from cluster 1 (overall lowest binding free energy), 4 and 6 (two largest clusters in the histogram) are shown. This is akin to the observations for docking to the protein surface excluding the binding site in FMN-bp and suggests that while heme 2 is the likely binding site for FMN, there is no single well-defined binding pose to heme 2.

The difference in FMN binding affinity between MtrC and the flavin-

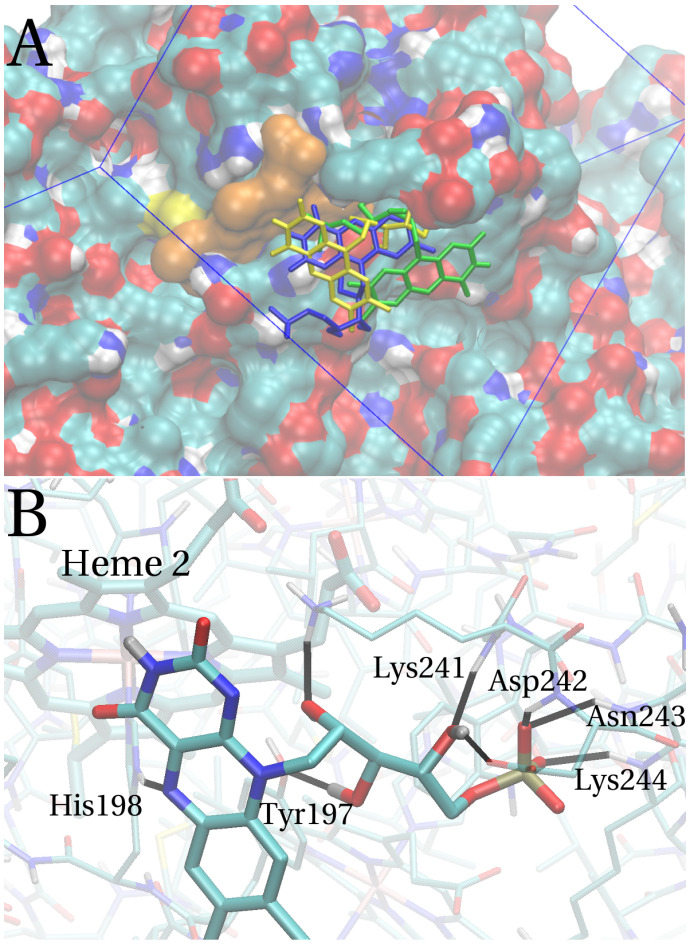

Figure 5.5: Docking of FMN to heme 2 in the crystal structure of MtrC. (A) Best poses of FMN in clusters 1 (yellow), 4 (green) and 6 (blue) of the histogram shown in Fig. 5.4 (i. e. the strongest-binding cluster and the two most populated clusters). Heme 2 is shown in orange and the rectangular Autodock search box in blue. (B) Close-up on best pose of FMN in cluster 1, indicating individual hydrogen bonds (black) together with the protein residues involved. The closest distance between the planar head group of FMN and the porphyrin edge is 3.5 Å.

binding proteins studied in section 5.2.1 can be explained qualitatively by the number of hydrogen bonds formed with the proteins (recognising, of course, that other interactions may be important as well). The best pose obtained by docking of FMN to heme 2 of MtrC forms 8 hydrogen bonds with the protein (Fig. 5.5B), compared to 11-13 hydrogen bonds formed with the flavin-binding proteins (Fig.5.3C,D). It is the head group that is hardly involved in bonding to MtrC (1 hydrogen bond), compared to the flavin-binding proteins (4 hydrogen bonds). These observations are consistent with the idea of the absence of a well-defined binding site in MtrC: The flexible ribitol side chain and phosphate tail can be expected to easily form hydrogen bonds on a globular protein surface while the requirements for the large rigid head group of FMN to form multiple hydrogen bonds (as in FMN-bp and FerB) is higher because it does not contain rotatable polar groups that could be easily positioned

to interact with H-bond acceptors or donors.

5.2.3 Structural flexibility of MtrC in the SH state

In experiment, FMN was found to bind much stronger to MtrC with the disulphide bond cleaved (SH state) [3]. Since no experimental structure is available for the SH state, it was attempted to rationalize this finding in this work by carrying out simulated annealing (SA) MD runs to probe the flexibility of the protein in this state. SA runs were carried out for both the SH and the SS state, which allowed us to relate any differences observed to the cleavage of the disulphide bond (see protocol). The front part of the loop containing the two cysteines (above heme 7; see Figure 1.5A) was found to show at least some flexibility in both sets of simulations; however, in the SS state this section can only flip upwards, whereas in the SH state the entire loop can lift off the barrel and move away even within the comparatively short SA protocol of a few ns.

Upon further MD simulation of one such structure for more than 100 ns at room temperature, the loop was found to actually move all the way to the front of the protein, somewhat inserting itself between Domain II to the left and some other loop of Domain IV to the right, seemingly forming a small hydrophobic core together with nonpolar residues from one loop in Domain II. The backbone seemed somewhat stable over the last few tens of ns, suggesting that this could possibly be a stable structure. This conformational switch is depicted in Fig. 5.6: Fig. 5.6A illustrates the motion experienced by the loop containing cysteine 453 by comparing the crystal structure of MtrC (blue) to the final structure after SA and further 110 ns simulation at room temperature (red). The final structure is presented in a space-filling representation in Panel B so as to show how the loop inserts itseld between Domains II and IV. Panel C depicts a new docking of FMN to MtrC after the observed conformational change (see next section).

Interestingly, the backbone of the barrel in Domain III itself is largely unaffected by the structural changes in the loop region during SA and even by the switching motion of the loop to the front shown in Fig. 5.6; in particular, no significant structural changes could be observed in the vicinity of heme 7 (which was hypothesised to be a possible binding site in the SH state[3]), despite the relatively high starting temperatures in the SA runs (600-700 K).

5.2.4 Docking to MtrC in the annealed SH state

As outlined in Section 5.1.4, dockings were also carried out several snapshots from the SA trajectories featuring the conformational switch of the cys-loop.

Five dockings were carried out at each of four potentially interesting regions. Features of interest in any poses obtained were both higher affinities and any qualitative binding site features that might suggest formation of some new binding site (even if no actually higher affinity could be found yet). Inspection of the results led to one potentially relevant binding pose shown in Fig. 5.6C with the flavin close to heme 4 and 5 (7.1 Å and 5.2 Å edge-to-edge distance, thus relevant for ET). While this pose does not yield an affinity better than 2 mM (weaker than the best pose at heme 2 in the SS state), it does show several features which are outlined here as they suggest the potential formation of an actual binding site.

The pose shown in Fig. 5.6C forms seven hydrogen bonds as well as an ionic bond. Three hydrogen bonds are formed with proprionates of heme 4 and 5 (left of the flavin) and four with residues from the cys-loop (right of the flavin), which after the conformational switch passes by in the vicinity of heme 4 (see also Fig. 5.6A). Two hydrogen bonds are formed by the flavin head group which also enters some kind of cleft, in contrast to the pose lying on the surface for heme 2 in the SS state. (The cluster histogram also showed a preference for the pose shown in Fig. 5.6C, unlike the unclear picture observed for heme 2 (see Figures 5.4 and 5.5).) Thus, different protein regions (the hemes 4 and 5 and the cys-loop) come together by the conformational switch to provide hydrogen bonds and ionic interactions for the flavin from both sides which might be the onset of an actual binding site. In the search for a binding site of FMN in the SH state that explains the strong affinity observed in experiment[3], this seems like the most interesting starting point. As mentioned in Section 5.1.4, there was also an interesting docking pose obtained in the general region around heme 7 - however, this was too far away to be relevant for ET.

5.3 Discussion

In this last part of our work we investigated docking of FMN to the solvent-accessible hemes in MtrC. At first we validated the docking protocol used by re-docking of FMN to flavin-binding proteins and found that both the experimental structure of the FMN-protein complex as well as a binding affinity could be reproduced in very good agreement with experiment. Using the same docking protocol we found that the interaction of FMN with MtrC is much weaker than for the flavin-binding proteins. There is no well-defined binding site, although our docking studies in the SS state indicate that interaction with heme 2 is strongest with a K_D of 490 μM, in close agreement with the experimental value of 255 μM.

A possible qualitative explanation for the relatively weak binding is the limited number of hydrogen bonds that the planar head group of FMN can form with MtrC. In the lowest free energy pose obtained, only one hydrogen bond is formed between the head group and MtrC, whereas 4 strong hydrogen bonds are formed with the FMN-binding proteins. Overall, the relatively weak and reversible binding of MtrC seems to be consistent with the role of a redox shuttle: binding to heme 2 in MtrC is stronger than on the protein surface but still relatively weak so that after ET rapid unbinding is possible. This presumes, however, that the interaction with the reduced forms of FMN is similarly weak, which we have not further investigated in this study.

Our results are consistent with previous docking studies on MtrC homologues, investigating the interaction of FMN with the deca-heme protein OmcA and of a related redox molecule, Anthraquinone 2,6-disulfonate, with the undeca-heme protein UndA [13]. In both cases heme 2 was identified as the preferred docking site, similar to the present result for MtrC. This begs the question whether there is any functional relevance for interaction with this particular heme. The microscopic redox potential computed for heme 2 in MtrF was at the higher end among the ten hemes, -0.06 V (see Section 3.1), implying that electron transfer to FMN ($\varepsilon^0 = -0.22$ V) is possible but would be slightly uphill if specific interactions between protein and FMN are neglected. Interestingly, the hemes that appeared as plausible candidates for shuttle binding sites due to their low (computed) redox potentials, heme 4 (-0.27 V) and heme 9 (-0.28 V), did not exhibit significant binding affinity for FMN.

Prompted by the recent suggestion that reduction of the disulphide bond in MtrC (SH state) strongly increases FMN binding, possible conformational switches upon cleavage of this bond were investigated via simulated annealing MD. A major conformational change of the loop containing the disulphide group upon cleavage of the bond was observed indeed, but a new binding site could not yet be clearly established. Nonetheless, it is interesting to note that the switch observed would suggest possible formation of a binding site somewhere else than in the barrel region next to heme 7, as was hypothesized previously[3]; in particular, the dockings in the SH state did yield a pose in the vicinity of heme 4 that for several reasons appears as a promising candidate for a new binding site.

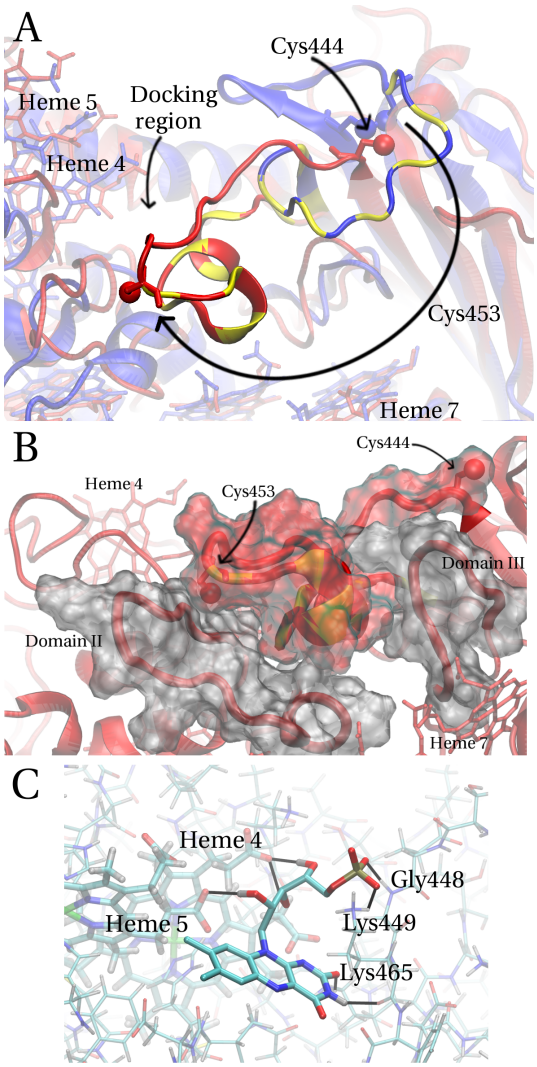

Figure 5.6: Simulated Annealing (SA) of MtrC in the SH state and subsequent docking of FMN. (A) The conformational switch of the cys-loop after SA and subsequent 110 ns of room temperature dynamics. The blue structure shows the crystal structure and the red structure depicts the final MD structure. The two cysteines are labelled and depicted explicitly in Licorice (as are the hemes), with sulphur as van der Waals spheres. For ease of comparison, the front part of the loop (according to the crystal structure position) is highlighted with yellow stripes in both structures. The long arrow for Cys453 describes its observed translocation. "Docking region" refers to the region where FMN was docked after SA (see panel C). (B) Close-up of the final position (red structure) of the cys-loop. The cysteines are labelled again and the yellow highlighting is kept. In addition, loops belonging to Domains II and III, respectively, are shown and labelled. These loops, as well as the cys-loop, are also shown in Surf representation to illustrate their spatial extension (red for the cys-loop, silver for the loops from Domains II and III). (C) Docking of FMN to the region around heme 4 after SA and the observed conformational switch. Hydrogen bonds are indicated as black lines. It can be seen that the flavin simultaneously interacts with the proprionates of heme 4 and 5 on the left side and with backbone and side chains of protein residues on the right side. These residues belong to the cys-loop, i. e. this binding pose is only possible after the conformational switch (see also position of the docking region in Panel (A)).

Chapter 6

Conclusion and Outlook

In this work, extensive computer simulations have provided thermodynamic and kinetic parameters for electron transfer in the bacterial deca-heme cytochrome MtrF, as well as insight into binding of a flavin to MtrC, a homologue of MtrF.

Our calculations of ET parameters in MtrF (specifically, redox potentials, reorganization free energies and electronic coupling matrix elements) enabled the calculation of heme-to-heme electron transfer rates in the framework of nonadiabatic Marcus theory, and they also allowed inferences regarding structure-function relationships. The redox potentials for a single excess electron in MtrF revealed, rather counter-intuitively, a rugged, symmetric free energy landscape featuring two energetic "hills" of around 0.2 eV for electron flow through the protein. One possible biological role of these low-potential hemes might be reduction of low-potential soluble redox shuttles like flavins, but thus far this is just a hypothesis. Electronic couplings were found to show the same symmetry along the cofactor network and also to correlate with redox potentials, with the highest electronic couplings occuring for the energetically most unfavourable ET steps. Modeling of through-protein electron transport showed that this correlation was necessary to maintain the steady-state electron transport rates observed in solution experiments, suggesting that this correlation is indeed relevant. Another interesting outcome of the simulations is that couplings and redox potentials can be modulated practically independently from each other, with implications for protein design. The electron transport rates obtained could however not reproduce the nA-currents observed in single-molecule I-V-measurements.

The protocols that were developed during the work for this thesis can also be transferred to other multi-heme cytochromes, namely MtrC, in order to obtain Marcus parameters for this homologue as well and see how they compare to the parameters obtained for MtrF. This is work currently underway by coworkers. The long-term goal for MtrC is comparison to experiment (in col-

laboration with experimentalists): Sophisticated site-directed mutations are planned to introduce an electron donor group as well as a differently ligated heme cofactor. According to this plan, excitation of the donor group should lead to controlled electron injection into a specific heme; the ligand-modified heme could then be spectroscopically distinguished from the other cofactors so as to spectroscopically follow the arrival of the electron at this reporter heme (starting from heme where the electron was injected), thus allowing to measure through-protein electron transport rates in experiment. These experimental electron transport rates could then be compared to rates obtained from simulation.

The disagreement between calculated electron transport rates assuming a heme-to-heme hopping model on the one hand and the experimentally observed currents in I-V-measurements on the other hand is still an unsolved puzzle. Thus, an alternative approach is currently being developed by coworkers where the electronic wave function would be explicitly propagated along the heme network in MtrF in the framework of nonadiabatic dynamics, thereby making no assumption regarding the validity of a hopping model.

Docking simulations of the small organic molecule FMN to MtrC are in agreement with the suggested role of FMN as a redox shuttle: The binding affinity of FMN to MtrC is significant but still weak, suggesting transient interactions. However, very recent experimental studies suggest the ability of MtrC to bind FMN much more stably if a disulphide bond is reduced and hence cleaved. Simulated Annealing studies carried out here yielded a major conformational change indeed upon cleavage of the disulphide bond, but the region where formation of a new binding site was suggested in experiment was found to be not visibly affected by the annealing. This suggested that the binding site might be formed somewhere else. Indeed, dockings of FMN to the annealed protein structure yielded a promising pose in the vicinity of heme 4, one of the hemes found to have a low redox potential in MtrF. A significant increase in affinity was not found; however, these dockings were more of an exploratory nature and one possible extension of this work could be a more extensive search for new binding sites as a consequence of the observed conformational change. In addition, this conformational switch in itself could be investigated more extensively; in a first step by qualitatively investigating whether long continuations of annealing-derived conformations usually yield this switch and, if the potential relevance of the switch observed seems high enough, by free energy calculations to estimate whether the new structure really is more stable than the initial structure (with the disulphide bond cleaved but no conformational change yet).

Overall, the work in this thesis has lead to a number of functional in-

sights into outer membrane deca-heme cytochromes, complementing experiment with molecular-level insight. Points of continuation that could be or are already being pursued are analogous calculations of Marcus parameters for homologues of MtrF (in connection with experimental studies); more extensive studies starting off from the state of the current work regarding conformational switches and flavin binding in the deca-heme cytochrome MtrC; and studies aiming at solving the discrepancies observed in this work between calculated and experimentally observed through-protein currents in I-V-measurements.

Bibliography

[1] M. Breuer, K. M. Rosso, J. Blumberger, and J. N. Butt. Multi-heme cytochromes in shewanella oneidensis mr-1: Structures, functions and opportuntities. *J. R. Soc. Interface*, 12:20141117, 2015.

[2] T.A. Clarke, M. Edwards, A. Gates, A. Hall, G. White, J. Bradley, C. Reardon, L. Shi, A. Beliaev, M. Marshall, Z. Wang, N. Watmough, J. Fredrickson, J. Zachara, J.N. Butt, and D.J. Richardson. Structure of a bacterial cell surface decaheme electron conduit. *Proc. Natl. Acad. Sci. USA*, 108:9384, 2011.

[3] Marcus J. Edwards, Gaye F. White, Michael Norman, Alice Tome-Fernandez, Emma Ainsworth, Liang Shi, Jim K. Fredrickson, John M. Zachara, Julea N. Butt, David J. Richardson, and Thomas A. Clarke. Redox Linked Flavin Sites in Extracellular Decaheme Proteins Involved in Microbe-Mineral Electron Transfer. *Scientific Reports*, 5, JUL 1 2015.

[4] M. Breuer, K. M. Rosso, and J. Blumberger. Electron flow in multi-heme bacterial cytochromes is a balancing act between heme electronic interaction and redox potentials. *Proc. Nat. Acad. Sci. USA*, 111:611, 2014.

[5] J. Blumberger. Free energies for biological electron transfer from qm/mm calculation: method, application and critical assessment. *Phys. Chem. Chem. Phys*, 10:5651, 2008.

[6] M. Breuer, P. Zarzycki, J. Blumberger, and K. M. Rosso. Thermodynamics of electron flow in the bacterial deca-heme cytochrome mtrf. *J. Am. Chem. Soc.*, 134:9868, 2012.

[7] M. Breuer, P. Zarzycki, L. Shi, T. A. Clarke, M. Edwards, J. Butt, D. J. Richardson, J. K. Fredrickson, J. M. Zachara, J. Blumberger, and K. M. Rosso. Molecular structure and free energy landscape for electron transport in the deca-heme cytochrome mtrf. *Biochem. Soc. Trans.*, 40:1198, 2012.

[8] D. M. A. Smith, K. M. Rosso, M. Dupuis, M. Valiev, and T. P. Straatsma. Electronic coupling between heme electron-transfer centers and its decay with distance depends strongly on relative orientation. *J. Phys. Chem. B*, 110:15582, 2006.

[9] V. Tipmanee and J. Blumberger. Kinetics of the terminal electron transfer step in cytochrome c oxidase. *J. Phys. Chem. B*, 116:1876, 2012.

[10] G. F. White, Z. Shi, L. Shi, Z. Wang, A. C. Dohnalkova, M. J. Marshall, J. K. Fredrickson, J. M. Zachara, J. N. Butt, D. J. Richardson, and T. A. Clarke. Rapid electron exchange between surface-exposed bacterial cytochromes and fe(iii) minerals. *Proc. Nat. Acad. Sci. USA*, 110:6346, 2013.

[11] V. Sedlacek, T. Klumpler, J. Marek, and I. Kucera. The structural and functional basis of catalysis mediated by nad(p)h:acceptor oxidoreductase (ferb) of paracoccus denitrificans. *PLOS ONE*, 9(5):e96262, 2014.

[12] E Liepinsh, M Kitamura, T Murakami, T Nakaya, and G Otting. Pathway of chymotrypsin evolution suggested by the structure of the fmn-binding protein from desulfovibrio vulgaris (miyazaki f). *Nat. Struct. Biol.*, 4:975, 1997.

[13] Catarina M. Paquete and Ricardo O. Louro. Unveiling the details of electron transfer in multicenter redox proteins. *Acc. Chem. Res.*, 47:56, 2014.

[14] H.B. Gray and J. R. Winkler. Electron flow through metalloproteins. *Biochim. Biophys. Acta*, 1797:1563, 2010.

[15] B. S. Brunschwig, C. Creutz, D. H. McCartney, T-K. Sham, and N. Sutin. The role of inner-sphere configuration changes in electron-exchange reactions of metal-complexes. *Faraday Discuss. Chem. Soc.*, 74:113, 1982.

[16] C. C. Moser, J. M. Keske, K. Warncke, R. S. Farid, and P. L. Dutton. Nature of biological electron-transfer. *Nature*, 355:796, 1992.

[17] H. B. Gray and J. R. Winkler. Electron tunneling through proteins. *Quart. Rev. Biophys.*, 36:341, 2003.

[18] D. J. Richardson, J. K. Fredrickson, and J. M. Zachara. Electron transport at the microbe-mineral interface: a synthesis of current research challenges. *Biochem. Soc. Trans.*, 40:1163, 2012.

[19] C. R. Myers and K. H. Nealson. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science*, 240:1319, 1988.

[20] R. S. Hartshorne, C. L. Reardon, D. Ross, J. Nuester, T. A. Clarke, A. J. Gates, P. C. Mills, J. K. Fredrickson, J. M. Zachara, L. Shi, A. S. Beliaev, M. J. Marshall, M. Tien, S. Brantley, J. N. Butt, and D. J. Richardson. Characterization of an extracellular conduit between bacteria and the extracellular environment. *Proc. Natl. Acad. Sci. USA*, 106:22169, 2009.

[21] Y. J. Xiong, L. Shi, B. W. Chen, M. U. Mayer, B. H. Lower, Y. Londer, S. Bose, M. F. Hochella, J. K. Fredrickson, and T. C. Squier. High-affinity binding and direct electron transfer to solid metals by the shewanella oneidensis mr-1 outer membrane c-type cytochrome omca. *J. Am. Chem. Soc.*, 128:13978, 2006.

[22] H. M. Jensen, A. E. Albers, K. R. Malley, Y. Y. Londer, B. E. Cohen, B. A. Helms, P. Weigele, J. T. Groves, and C. M. Ajo-Franklin. Engineering of a synthetic electron circuit in living cells. *Proc. Natl. Acad. Sci. USA*, 107:19213, 2010.

[23] K. Venkateswaran, D.P. Moser, M.E. Dollhopf, D.P. Lies, D.A. Saffarini, B.J. MacGregor, D.B. Ringelberg, D.C. White, M. Nishijima, H. Sano, et al. Polyphasic taxonomy of the genus Shewanella and description of Shewanella oneidensis sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 49(2):705, 1999.

[24] L. J. Smith, A. Kahraman, and J. M. Thornton. Heme proteins-diversity in structural characteristics, function, and folding. *Proteins-Structure Function and Bioinformatics*, 78(10):2349–2368, 2010.

[25] A. K. Sharma, G. Cavallaro, and A. Rosato. A systematic investigation of multiheme c-type cytochromes in prokaryotes. *J. Biol. Inorg. Chem.*, 15:559, 2010.

[26] S. Kerisit, K. M. Rosso, M. Dupuis, and M. Valiev. Molecular computational investigation of electron-transfer kinetics across cytochrome-iron oxide interfaces. *J. Phys. Chem. C*, 111:11363, 2007.

[27] E. D. Brutinel and J. A. Gralnick. Shuttling happens: soluble flavin mediators of extracellular electron transfer in shewanella. *Appl. Microbiol. Biotechnol.*, 93:41, 2012.

[28] D. Coursolle, D. B. Baron, D. R. Bond, and J. A. Gralnick. The mtr respiratory pathway is essential for reducing flavins and electrodes in shewanella oneidensis. *Journal of Bacteriology*, 192(2):467–474, 2010.

[29] E. Marsili, D. B. Baron, I. D. Shikhare, D. Coursolle, J. A. Gralnick, and D. R. Bond. Shewanella secretes flavins that mediate extracellular electron transfer. *Proc. Natl. Acad. Sci. USA*, 105:3968, 2008.

[30] D.P. Lies, M.E. Hernandez, A. Kappler, R.E. Mielke, J.A. Gralnick, and D.K. Newman. Shewanella oneidensis MR-1 uses overlapping pathways for iron reduction at a distance and by direct contact under conditions relevant for biofilms. *Applied and environmental biology*, 71(8):4414, AUG 2005.

[31] Y.A. Gorby, S. Yanina, J.S. McLean, K.M. Rosso, D. Moyles, A. Dohnalkova, T.J. Beveridge, I.S. Chang, B.H. Kim, K.S. Kim, D.E. Culley, S.B. Reed, M.F. Romine, D.A. Saffarini, E.A. Hill, L. Shi, D.A. Elias, D.W. Kennedy, G. Pinchuk, K. Watanabe, S. Ishii, B. Logan, K. H. Nealson, and J. K. Fredrickson. Electrically conductive bacterial nanowires produced by shewanella oneidensis strain mr-1 and other microorganisms. *Proc. Nat. Acad. Sci. USA*, 103:11358, 2006.

[32] M. Y. El-Naggar, G. Wanger, K.M. Leung, T. D. Yuzvinsky, G. Southam, J. Yang, W. M. Lau, K. H. Nealson, and Y. A. Gorby. Electrical transport along bacterial nanowires from shewanella oneidensis mr-1. *Proc. Natl. Acad. Sci. USA*, 107:18127, 2010.

[33] K. M. Leung, G. Wanger, M. Y. El-Naggar, Y. A. Gorby, G. Southam, W. M. Lau, and J. Yang. Shewanella oneidensis mr-1 bacterial nanowires exhibit p-type, tunable electronic behavior. *ACS Nano*, 13:2407, 2013.

[34] S. Pirpadian, S. E. Barchinger, K. M. Leung, H. S. ByunY. Jangir, R. A. Bouhenni, S. B. Reed, M. F. Romine, D. A. Saffarini, L. Shi, Y. A. Gorby, J. H. Golbeck, and M. Y. El-Naggar. Shewanella oneidensis mr-1 nanowires are outer membrane and periplasmic extensions of the extracellular electron transport components. *Proc. Natl. Acad. Sci.*, 111:12883, 2014.

[35] A. Okamotoa, K. Hashimotoa, K. H. Nealson, and R. Nakamura. Rate enhancement of bacterial extracellular electron transport involves bound flavin semiquinones. *Proc. Nat. Acad. Sci. USA*, 110:7856, 2013.

[36] L. Shi, B. Chen, Z. Wang, D.A. Elias, M.U. Mayer, Y.A. Gorby, S. Ni, B.H. Lower, D.W. Kennedy, D.S. Wunschel, et al. Isolation of a high-affinity functional protein complex between OmcA and MtrC: two outer membrane decaheme c-type cytochromes of Shewanella oneidensis MR-1. *Journal of Bacteriology*, 188(13):4705, 2006.

[37] D. E. Ross, S. S. Ruebush, S. L. Brantley, R. S. Hartshorne, T. A. Clarke, D. J. Richardson, and M. Tien. Characterization of protein-protein interactions involved in iron reduction by shewanella oneidensis mr-1. *Applied and Environmental Microbiology*, 73(18):5797–5808, 2007.

[38] H. Z. Zhang, X. T. Tang, G. R. Munske, N. Zakharova, L. Yang, C. X. Zheng, M. A. Wolff, N. Tolic, G. A. Anderson, L. Shi, M. J. Marshall, J. K. Fredrickson, and J. E. Bruce. In vivo identification of the outer membrane protein omca-mtrc interaction network in shewanella oneidensis mr-1 cells using novel hydrophobic chemical cross-linkers. *Journal of Proteome Research*, 7(4):1712–1720, 2008.

[39] D. Coursolle and J. A. Gralnick. Modularity of the mtr respiratory pathway of shewanella oneidensis strain mr-1. *Mol. Microbiol.*, 77:995, 2010.

[40] J. S. McLean, G. E. Pinchuk, O. V. Geydebrekht, C. L. Bilskis, B. A. Zakrajsek, E. A. Hill, D. A. Saffarini, M. F. Romine, Y. A. Gorby, J. K. Fredrickson, and A. S. Beliaev. Oxygen-dependent autoaggregation in shewanella oneidensis mr-1. *Environmental Microbiology*, 10(7):1861–1876, 2008.

[41] D. G. G. McMillan, S. J. Marritt, J. N. Butt, and L. J. C. Jeuken. Menaquinone-7 is specific cofactor in tetraheme quinol dehydrogenase cyma. *Journal of Biological Chemistry*, 287(17):14215–14225, 2012.

[42] B.H. Lower, L. Shi, R. Yongsunthon, T.C. Droubay, D.E. McCready, and S.K. Lower. Specific bonds between an iron oxide surface and outer membrane cytochromes MtrC and OmcA from Shewanella oneidensis MR-1. *Journal of Bacteriology*, 189(13):4944, 2007.

[43] M. J. Edwards, A. Hall, L. Shi, J. K. Fredrickson, J. M. Zachara, J. N. Butt, D. J. Richardson, and T. A. Clarke. The crystal structure of the extracellular 11-heme cytochrome unda reveals a conserved 10-heme motif and defined binding site for soluble iron chelates. *Structure*, 20:1275, 2012.

[44] Bruno M. Fonseca, Catarina M. Paquete, Sania E. Neto, Isabel Pacheco, Claudio M. Soares, and Ricardo O. Louro. Mind the gap: cytochrome interactions reveal electron pathways across the periplasm of shewanella oneidensis mr-1. *Biochemical Journal*, 449(1):101–108, 2013.

[45] R. S. Hartshorne, B. N. Jepson, T. A. Clarke, S. J. Field, J. Fredrickson, J. Zachara, L. Shi, J. N. Butt, and D. J. Richardson. Characterization of shewanella oneidensis mtrc: a cell-surface decaheme cytochrome involved in respiratory electron transport to extracellular electron acceptors. *J. Biol. Inorg. Chem.*, 12:1083, 2007.

[46] Marcus J. Edwards, Nanakow A. Baiden, Alexander Johs, Stephen J. Tomanicek, Liyuan Liang, Liang Shi, Jim K. Fredrickson, John M. Zachara, Andrew J. Gates, Julea N. Butt, David J. Richardson, and Thomas A. Clarke. The x-ray crystal structure of shewanella oneidensis omca reveals new insight at the microbe-mineral interface. *FEBS Letters*, 588(10):1886 – 1890, 2014.

[47] Dayle M. A. Smith and Kevin M. Rosso. Possible Dynamically Gated Conductance along Heme Wires in Bacterial Multiheme Cytochromes. *J. Phys. Chem. B*, 118(29):8505–8512, JUL 24 2014.

[48] Hye Suk Byun, Sahand Pirbadian, Aiichiro Nakano, Liang Shi, and Mohamed Y. El-Naggar. Kinetic Monte Carlo Simulations and Molecular Conductance Measurements of the Bacterial Decaheme Cytochrome MtrF. *CHEMELECTROCHEM*, 1(11, SI):1932–1939, NOV 11 2014.

[49] C. Masato Nakano, Hye Suk Byun, Heng Ma, Tao Wei, and Mohamed Y. El-Naggar. A framework for stochastic simulations and visualization of biological electron-transfer dynamics. *Computer Physics Communications*, 193:1 – 9, 2015.

[50] R. A. Marcus. Theory of oxidation-reduction reactions involving electron transfer .1. *J. Chem. Phys.*, 24:966, 1956.

[51] R. A. Marcus and N. Sutin. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta*, 811:265, 1985.

[52] T. Van Voorhis, T. Kowalczyk, B. Kaduk, L.-P. Wang, C.-L. Cheng, and Q. Wu. *Ann. Rev. Phys. Chem.*, 61:149, 2010.

[53] W. F. Libby. Theory of electron exchange reactions in aqueous solution. *Journal of Physical Chemistry*, 56(7):863–868, 1952.

[54] J. K. Hwang and A. Warshel. Microscopic examination of free-energy relationships for electron transfer in polar solvents. *J. Am. Chem. Soc.*, 109:715, 1987.

[55] C. C. Moser, C. C. Page, and P. L. Dutton. Darwin at the molecular scale: selection and variance in electron tunnelling proteins including cytochrome c oxidase. *Phil. Trans. R. Soc. B*, 361:1295, 2006.

[56] C. C. Moser, C. C. Page, X. Chen, and P. L. Dutton. Biological electron tunneling through native protein media. *J. Biol. Inorg. Chem.*, 2:393, 1997.

[57] Andrew R. Leach. *Molecular modelling: principles and applications*, page 126ff. Addison-Wesley Longman Ltd, 2001.

[58] Piotr Cieplak, Francois-Yves Dupradeau, Yong Duan, and Junmei Wang. Polarization effects in molecular mechanical force fields. *JOURNAL OF PHYSICS-CONDENSED MATTER*, 21(33), 2009.

[59] Andrew R. Leach. *Molecular modelling: principles and applications*, page 228. Addison-Wesley Longman Ltd, 2001.

[60] Andrew R. Leach. *Molecular modelling: principles and applications*, page 165ff. Addison-Wesley Longman Ltd, 2001.

[61] Andrew R. Leach. *Molecular modelling: principles and applications*, page 317ff. Addison-Wesley Longman Ltd, 2001.

[62] P. P. Ewald. *Ann. Phys.*, 64:253, 1921.

[63] S. W. deLeeuw, J. W. Perram, and E. R. Smith. *Proc. Roy. Soc. A*, 373:27, 1980.

[64] Andrew R. Leach. *Molecular modelling: principles and applications*, page 324. Addison-Wesley Longman Ltd, 2001.

[65] Andrew R. Leach. *Molecular modelling: principles and applications*, page 353ff. Addison-Wesley Longman Ltd, 2001.

[66] Andrew R. Leach. *Molecular modelling: principles and applications*, page 410ff. Addison-Wesley Longman Ltd, 2001.

[67] Andrew R. Leach. *Molecular modelling: principles and applications*, page 423. Addison-Wesley Longman Ltd, 2001.

[68] Andrew R. Leach. *Molecular modelling: principles and applications*, page 357. Addison-Wesley Longman Ltd, 2001.

[69] M.P. Allen and D.J. Tildesley. *Computer simulation of liquids*. Clarendon Press, 1999.

[70] Andrew R. Leach. *Molecular modelling: principles and applications*, page 630f. Addison-Wesley Longman Ltd, 2001.

[71] V. Tipmanee, H. Oberhofer, M. Park, K. S. Kim, and J. Blumberger. Prediction of reorganization free energies for biological electron transfer: A comparative study of ru-modified cytochromes and a 4-helix bundle protein. *J. Am. Chem. Soc.*, 132:17032, 2010.

[72] K. Izutsu. *Electrochemistry in Nonaqueous Solutions*. VCH, Weinheim, 2002.

[73] Kieron Burke. Perspective on density functional theory. *JOURNAL OF CHEMICAL PHYSICS*, 136(15), 2012.

[74] P. Hohenberg and W. Kohn. *Phys. Rev. B*, 136:864, 1964.

[75] W. Kohn and L. J. Sham. *Phys. Rev. A*, 140:1133, 1965.

[76] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.

[77] B. Miehlich, A. Savin, H. Stoll, and H. Preuss. *Chem. Phys. Lett.*, 157:200, 1989.

[78] K. Kim and K. D. Jordan. Comparison of density functional and mp2 calculations on the water monomer and dimer. *The Journal of Physical Chemistry*, 98(40):10089–10094, 1994.

[79] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry*, 98(45):11623–11627, 1994.

[80] Dominik Marx and Juerg Hutter. *Ab initio molecular dynamics*, page 85ff. Cambridge University Press, 2009.

[81] Dominik Marx and Juerg Hutter. *Ab initio molecular dynamics*, page 267ff. Cambridge University Press, 2009.

[82] H. Oberhofer and J. Blumberger. Electronic coupling matrix elements from charge constrained density functional theory calculations using a plane wave basis set. *J. Chem. Phys.*, 133:244105, 2010.

[83] M. D. Newton. Quantum chemical probes of electron-transfer kinetics: The nature of donor-acceptor interactions. *Chem. Rev.*, 91:767, 1991.

[84] Q. Wu and T. Van Voorhis. *J. Chem. Phys.*, 125:164105, 2006.

[85] R ROSENFELD, S VAJDA, and C DELISI. FLEXIBLE DOCKING AND DESIGN. *ANNUAL REVIEW OF BIOPHYSICS AND BIOMOLECULAR STRUCTURE*, 24:677–700, 1995.

[86] GM Morris, DS Goodsell, RS Halliday, R Huey, WE Hart, RK Belew, and AJ Olson. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *JOURNAL OF COMPUTATIONAL CHEMISTRY*, 19(14):1639–1662, NOV 15 1998.

[87] Garrett M. Morris, Ruth Huey, William Lindstrom, Michel F. Sanner, Richard K. Belew, David S. Goodsell, and Arthur J. Olson. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *J. Comp. Chem.*, 30:2785, 2009.

[88] J. H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.

[89] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. A. Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comp. Chem.*, 24:1999, 2003.

[90] M.C. Lee and Y. Duan. Distinguish protein decoys by using a scoring function based on a new amber force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins: Structure, Function, and Bioinformatics*, 55(3):620, 2004.

[91] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926, 1983.

[92] I.S. Joung and T.E. Cheatham III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112:9020, 2008.

[93] C. R. Sondergaard, M. H. M. Olsson, M. Rostkowski, and J. H. Jensen. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pka values. *J. Chem. Theory Comput.*, 7:2284, 2011.

[94] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M. G. Seetin, C. Sagui, V. Babin, and P.A. Kollman. AMBER 10, University of California, San Francisco, 2008.

[95] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with namd. *J. Comput. Chem.*, 26:1781, 2005.

[96] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33, 1996.

[97] Kendall Atkinson. *An introduction to numerical analysis*. Wiley, New York, 1989.

[98] P Cieplak, J Caldwell, and P Kollman. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: Aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *JOURNAL OF COMPUTATIONAL CHEMISTRY*, 22(10):1048–1057, 2001.

[99] ZX Wang, W Zhang, C Wu, HX Lei, P Cieplak, and Y Duan. Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *JOURNAL OF COMPUTATIONAL CHEMISTRY*, 27(6):781–790, 2006.

[100] J. W. Caldwell and P. A. Kollman. Structure and properties of neat liquids using nonadditive molecular-dynamics - water, methanol, and n-methylacetamide. *J. Phys. Chem.*, 99:6208, 1995.

[101] Joseph P Brown. Reduction of polymeric azo and nitro dyes by intestinal bacteria. *Applied and environmental microbiology*, 41(5):1283–1286, 1981.

[102] CPMD, http://www.cpmd.org . *Copyright IBM Corp 1990-2008, Copyright MPI fuer Festkoerperforschung Stuttgart 1997-2001* .

[103] Nieplocha, E. Apra, T.L. Windus, M. Valiev W.A. de Jong, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. Van Dam, D. Wang, and J. Nwchem: A comprehensive and scalable open-source

solution for large scale molecular simulations. *Comp. Phys. Comm.*, 181:1477, 2010.

[104] C. Hartwigsen, S. Goedecker, and J. Hutter. Relativistic separable dual-space gaussian pseudopotentials from h to rn. *Phys. Rev. B*, 58:3641, 1998.

[105] K Akasaka. Probing conformational fluctuation of proteins by pressure perturbation. *CHEMICAL REVIEWS*, 106(5):1814–1835, MAY 2006.

[106] E Paci. High pressure simulations of biomolecules. *BIOCHIMICA ET BIOPHYSICA ACTA-PROTEIN STRUCTURE AND MOLECULAR ENZYMOLOGY*, 1595(1-2):185–200, MAR 25 2002.

[107] Michael T. Lerch, Zhongyu Yang, Evan K. Brooks, and Wayne L. Hubbell. Mapping protein conformational heterogeneity under pressure with site-directed spin labeling and double electron-electron resonance. *Proc. Natl. Acad. Sci. USA*, 111(13):E1201–E1210, APR 1 2014.

[108] A. Laio and M. Parrinello. *Proc. Natl. Acad. Sci. USA*, 20:12562, 2002.

[109] F. A. Walker. Magnetic spectroscopic (epr, eseem, mossbauer, mcd and nmr) studies of low-spin ferriheme centers and their corresponding heme proteins. *Coord. Chem. Rev.*, 185:471, 1999.

[110] C. C. Moser, S. E. Chobot, C. C. Page, and P. Leslie Dutton. Distance metrics for heme protein electron tunneling. *Biochim. Biophys. Acta*, 1777:1032, 2008.

[111] N. F. Polizzi, S. S. Skourtis, and D. N. Beratan. Physical constraints on charge transport through bacterial nanowires. *Faraday Discuss.*, 155:43, 2012.

[112] Sarah M. Strycharz-Glaven, Rachel M. Snider, Anthony Guiseppi-Elie, and Leonard M. Tender. On the electrical conductivity of microbial nanowires and biofilms. *Energy Environ. Sci.*, 4:4366, 2011.

[113] S. Pirbadian and M. Y. El-Naggar. Multistep hopping and extracellular charge transfer in microbial redox chains. *Phys. Chem. Chem. Phys.*, 14:13802, 2012.

[114] MATLAB. *version 7.11.0.584 (R2010b)*. The MathWorks Inc., Natick, Massachusetts, USA, 2010.

[115] KE Pitts, PS Dobbin, F Reyes-Ramirez, AJ Thomson, DJ Richardson, and HE Seward. Characterization of the shewanella oneidensis mr-1 decaheme cytochrome mtra. *J. Biol. Chem.*, 278:27758, 2003.

[116] T. A. Clarke, T. Holley, R. S. Hartshorne, J. K. Fredrickson, J. M. Zachara, L. Shi, and D. J. Richardson. The role of multihaem cytochromes in the respiration of nitrite in escherichia coli and fe(iii) in shewanella oneidensis. *Biochem. Soc. Trans.*, 36:1005, 2008.

[117] Mackenzie A. Firer-Sherwood, Nozomi Ando, Catherine L. Drennan, and Sean J. Elliott. Solution-based structural analysis of the decaheme cytochrome, mtra, by small-angle x-ray scattering and analytical ultracentrifugation. *J. Phys. Chem. B*, 115:11208, 2011.

[118] T. A. Clarke, J. A. Cole, D. J. Richardson, and A. M. Hemmings. The crystal structure of the pentahaem c-type cytochrome nrfb and characterization of its solution-state interaction with the penthaem nitrite reductase nrfa. *Biochem. J.*, 406:19, 2007.

[119] C. E. D. Chidsey. Free-energy and temperature-dependence of electron-transfer at the metal-electrolyte interface. *Science*, 251:919, 1991.

[120] C. A. Bortolotti, M. E. Siwko, E. Castellini, A. Ranieri, M. Sola, and S. Corni. The reorganization energy in cytochrome c is controlled by the accessibility of the heme to the solvent. *J. Phys. Chem. Lett.*, 2:1761, 2011.

[121] D.E. Ross, S.L. Brantley, and M. Tien. Kinetic characterization of OmcA and MtrC, terminal reductases involved in respiratory electron transfer for dissimilatory iron reduction in Shewanella oneidensis MR-1. *Applied and Environmental Microbiology*, 75(16):5218, 2009.

[122] C. Hetenyi and D. van der Spoel. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci*, 11:1729, 2002.

[123] C. Hetenyi and D. van der Spoel. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Letters*, 580:1447, 2006.

[124] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The resp model. *J. Phys. Chem.*, 97:10269, 1993.

[125] A. Klamt and G. Schuurmann. Cosmo: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2*, pages 799–805, 1993.